

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

A1

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification</b> <sup>6</sup> : <b>C12N 15/12, C07K 14/705, C12N 5/10, 15/57, 9/48, 9/14, 15/55</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 98/21328</b> <b>(43) International Publication Date:</b> 22 May 1998 (22.05.98)
<b>(21) International Application Number:</b> PCT/JP97/04056 <b>(22) International Filing Date:</b> 7 November 1997 (07.11.97)  <b>(30) Priority Data:</b> 8/301429 13 November 1996 (13.11.96) JP  <b>(71) Applicants (for all designated States except US):</b> SAGAMI CHEMICAL RESEARCH CENTER [JP/JP]; 4-1, Nishi-Ohnuma 4-chome, Sagamihara-shi, Kanagawa 229 (JP). PROTEGENE INC. [JP/JP]; 2-20-3, Naka-cho, Meguro-ku, Tokyo 153 (JP).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> KATO, Seishi [JP/JP]; 3-46-50, Wakamatsu, Sagamihara-shi, Kanagawa 229 (JP). SEKINE, Shingo [JP/JP]; 4-4-1, Nishi-Ohnuma, Sagami-hara-shi, Kanagawa 229 (JP). YAMAGUCHI, Tomoko [JP/JP]; 5-13-11, Takasago, Katsushika-ku, Tokyo 125 (JP). KOBAYASHI, Midori [JP/JP]; 647-2, Chougo, Fujisawa-shi, Kanagawa 252 (JP).  <b>(74) Agents:</b> AOYAMA, Tamotsu et al.; Aoyama & Partners, IMP Building, 3-7, Shiromi 1-chome, Chuo-ku, Osaka-shi, Osaka 540 (JP).	<b>(81) Designated States:</b> AU, CA, JP, MX, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>	
<b>(54) Title:</b> HUMAN PROTEINS HAVING TRANSMEMBRANE DOMAINS AND DNAS ENCODING THESE PROTEINS  <b>(57) Abstract</b>  Proteins containing any of the amino acid sequences represented by Sequence No. 1 to Sequence No. 2 or by Sequence No. 4 to Sequence No. 25 and DNAs encoding said proteins exemplified by cDNAs containing any of the base sequences represented by Sequence No. 26 to Sequence No. 50. Said proteins can be provided by expressing cDNAs encoding human proteins having transmembrane domains and recombinants of these human cDNAs.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## DESCRIPTION

Human Proteins Having Transmembrane  
Domains and DNAs Encoding These Proteins

## TECHINICAL FIELD

The present invention relates to human proteins having transmembrane domains, DNAs encoding these proteins and eukaryotic cells expressing those DNAs. The proteins of the present invention can be used as pharmaceuticals or as antigens for preparing antibodies against said proteins. The cDNAs of the present invention can be used as probes for the gene diagnosis and gene sources for the gene therapy. Furthermore, the cDNAs can be used as gene sources for large-scale production of the proteins encoded by said cDNAs. Moreover, the cells introduced with DNAs encoding transmembrane proteins therein and expressing transmembrane proteins in large amounts can be used for detection of the corresponding ligands as well as screening of novel low molecular medicines.

## BACKGROUND ART

Membrane proteins play important roles, as signal receptors, ion channels, transporters, etc., for the material transportation and the information transmission which are mediated by the cell membrane. Their examples include receptors for a variety of cytokines, ion channels for the sodium ion, the potassium ion, the chloride ion, etc., transporters for saccharides and amino acids, and so on,

where the genes for many of them have been cloned already.

It has been clarified that the abnormalities of these membrane proteins are related to a number of hitherto cryptogenic diseases. For example, a gene for a membrane protein having 12 transmembrane domains was identified as the gene responsible for cystic fibrosis [Rommens, J. M. et al., Science 245: 1059-1065 (1989)]. In addition, it has been clarified that several membrane proteins act as the receptors when a virus infects the cells. For example, HIV-1 is revealed to infect into the cells through the mediation of a membrane protein fusin, a membrane protein on the T-cell membrane, having a CD-4 antigen and 7 transmembrane domains [Feng, Y. et al., Science 272: 872-877 (1996)]. Therefore, discovery of a new membrane protein is anticipated to lead to the elucidation of the causes of many diseases, whereby isolation of a new gene coding for the membrane protein has been desired.

Heretofore, owing to difficulty in the purification, many of membrane proteins have been isolated by an approach from the gene side. A general method is the so-called expression cloning which comprises transfection of a cDNA library in the animal cells to express the cDNA and then detection of the cells expressing the target membrane protein on the membrane by an immunological technique using an antibody or a biological technique for the change in the membrane permeability. However, this method is applicable only to cloning of a gene for a membrane protein with a known function.

In general, membrane proteins possess hydrophobic



transmembrane domains inside the proteins which are synthesized in the ribosome and then remain in the phospholipid to be trapped in the membrane. Accordingly, the evidence of the cDNA for encoding the membrane protein is provided by determination of the whole base sequence of a full-length cDNA followed by detection of highly hydrophobic transmembrane domains in the amino acid sequence of the protein encoded by said cDNA.

The object of the present invention is to provide novel human proteins having transmembrane domains, DNAs encoding said proteins and transformed eukaryotic cells capable of expressing said DNAs.

As the result of intensive studies, the present inventors were successful in cloning of cDNAs having transmembrane domains from a human full-length cDNA bank, thereby completing the present invention. That is to say, the present invention provides proteins containing any of the amino acid sequences represented by Sequence No. 1 to Sequence No. 2 or by Sequence No. 4 to Sequence No. 25 that are human proteins having transmembrane domains. The present invention also provides DNAs encoding said proteins such as cDNAs containing any of the base sequences represented by Sequence No. 26 to Sequence No. 50 and transformed eukaryotic cells capable of expressing said DNAs.

Each of the proteins of the present invention can be obtained, for example, by a method for isolation from human organs, cell lines, etc, a method for preparation of the peptide by the chemical synthesis on the basis of the amino acid sequence of the present invention, or a method for

production with the recombinant DNA technology using the DNA encoding the transmembrane domains of the present invention, wherein the method for obtainment by the recombinant DNA technology is employed preferably. For example, an in vitro expression can be achieved by preparation of an RNA by the in vitro transcription from a vector having a cDNA of the present invention, followed by the in vitro translation using this RNA as a template. Also, the recombination of the translation domain to a suitable expression vector by the method known in the art leads to the expression of a large amount of the encoded protein by using prokaryotic cells (e.g. *Escherichia coli*, *Bacillus subtilis*) or eukaryotic cells (e.g. yeasts, insect cells, animal cells).

In the case in which a protein of the present invention is expressed by a microorganism such as *Escherichia coli*, the translation region of a cDNA of the present invention is constructed in an expression vector having an origin, a promoter, ribosome-binding site(s), cDNA-cloning site(s), a terminator, etc. that can be replicated in the microorganism and, after transformation of the host cells with said expression vector, the thus-obtained transformant is incubated, whereby the protein encoded by said cDNA can be produced on a large scale in the microorganism. In that case, a protein fragment containing an optional region can be obtained by performing the expression with inserting an initiation codon and a termination codon before and after the optional translation region. Alternatively, a fusion protein with another protein can be expressed. Only a protein portion encoding said cDNA can be obtained by cleavage of said fusion

protein with an appropriate protease.

In the case wherein a protein of the present invention is to be produced in eukaryotic cells, the translation region of said cDNA may be subjected to recombination to an expression vector for eukaryotic cells having a promoter, a splicing domain, a poly(A) addition site, etc. and transfected into the eukaryotic cells so that the protein is produced as a membrane protein on the cell membrane surface. As the expression vector, there are exemplified pKA1, pCDM8, pSVK3, pMSG, pSVL, pBK-CMV, pBK-RSV, EBV vector, pRS, pYES2, etc. Examples of the eukaryotic cells are mammalian animal culture cells (e.g. simian renal cells COS7, chinese hamster ovarian cells CHO), blast yeasts, fission yeasts, silkworm yeasts, South African clawed toad oocytes, etc. However, any eukaryotic cells may be used insofar as the protein of the invention can be expressed on the cell membrane surface. In order to introduce the expression vector into the eukaryotic cells, there may be used any per se conventional method such as electroporation method, calcium phosphate method, liposome method or DEAE dextran method.

For separation and purification of the protein of the invention from the culture after expression of such protein in prokaryotic cells or eukaryotic cells, conventional separation operations may be adopted, if necessary, in their proper combination. Examples of the conventional separation operations are treatment with a denaturing agent (e.g. urea) or a surfactant, ultrasonic treatment, enzymatic digestion, salting out, solvent precipitation, dialysis, centrifugation, ultrafiltration, gel filtration, SDS-PAGE, isoelectric point

electrophoresis, ion exchange chromatography, hydrophobic chromatography, affinity chromatography, reverse phase chromatography, etc.

The proteins of the present invention include peptide fragments (more than 5 amino acid residues) containing any partial amino acid sequence of the amino acid sequences represented by Sequence No. 1 to Sequence No. 2 or by Sequence No. 4 to Sequence No. 25. These fragments can be used as antigens for preparation of the antibodies. Also, the proteins of the present invention that have signal sequences appear in the form of maturation proteins on the cell surface, after the signal sequences are removed. Therefore, these maturation proteins shall come within the scope of the present invention. The N-terminal amino acid sequences of the maturation proteins can be easily identified by using the method for the cleavage-site determination in a signal sequence [Japanese Patent Kokai Publication No. 1996-187100]. Furthermore, many membrane proteins are subjected to the processing on the cell surface to be converted to the secretor forms. These secretor proteins or peptides shall come within the scope of the present invention. When glycosylation sites are present in the amino acid sequences, expression in appropriate animal cells affords glycosylated proteins. Therefore, these glycosylated proteins or peptides also shall come within the scope of the present invention.

The DNAs of the present invention include all DNAs encoding the above-mentioned proteins. Said DNAs can be obtained using the method by chemical synthesis, the method by cDNA cloning, and so on.

Each of the cDNAs of the present invention can be cloned from, for example, a cDNA library of the human cell origin. The cDNA is synthesized using as a template a poly(A)<sup>+</sup> RNA extracted from human cells. The human cells may be cells delivered from the human body, for example, by the operation or may be the culture cells. The cDNA can be synthesized by using any method selected from the Okayama-Berg method [Okayama, H. and Berg, P., Mol. Cell. Biol. 2: 161-170 (1982)], the Gubler-Hoffman method [Gubler, U. and Hoffman, J. Gene 25: 263-269 (1983)], and so on, but it is preferred to use the capping method [Kato, S. et al., Gene 150: 243-250 (1994)] as illustrated in Examples in order to obtain a full-length clone in an effective manner.

The primary selection of a cDNA encoding a human protein having transmembrane domain(s) is performed by the sequencing of a partial base sequence of the cDNA clone selected at random from the cDNA library, sequencing of the amino acid sequence encoded by the base sequence, and recognition of the presence or absence of hydrophobic site(s) in the resulting N-terminal amino acid sequence region. Next, the secondary selection is carried out by determination of the whole base sequence by the sequencing and the protein expression by the in vitro translation. The ascertainment of the cDNA of the present invention for encoding the protein having the secretory signal sequence is performed by using the signal sequence detection method [Yokoyama-Kobayashi, M. et al., Gene 163: 193-196 (1995)]. In other words, the ascertainment for the coding portion of the inserted cDNA fragment to function as a signal sequence is provided by

fusing a cDNA fragment encoding the N-terminus of the target protein with a cDNA encoding the protease domain of urokinase and then expressing the resulting cDNA in COS7 cells to detect the urokinase activity in the cell culture medium. On the other hand, the N-terminal region is judged to remain in the membrane in the case where the urokinase activity is not detected in the cell culture medium.

The cDNAs of the present invention are characterized by containing any of the base sequences represented by Sequence No. 26 to Sequence No. 50 and any of the base sequences represented by Sequence No. 51 to Sequence No. 75. Table 1 summarizes the clone number (HP number), the cells affording the cDNA, the total base number of the cDNA, and the number of the amino acid residues of the encoded protein, for each of the cDNAs.

Table 1

Sequence Number	HP Number	Cells	Number of Bases	Number of Amino Acid Residues
1, 26, 51	HP00442	HT-1080	986	205
2, 27, 52	HP00804	Leucocyte	1824	371
3, 28, 53	HP01098	Stomach cancer	1076	179
4, 29, 54	HP01148	Liver	1591	347
5, 30, 55	HP01293	Liver	1888	554
6, 31, 56	HP10013	KB	2033	350
7, 32, 57	HP10034	HT-1080	911	209
8, 33, 58	HP10050	HT-1080	601	163

9, 34, 59	HP10071	Stomach cancer	394	92
10, 35, 60	HP10076	U937	732	172
11, 36, 61	HP10085	U937	697	149
12, 37, 62	HP10122	Stomach cancer	1186	188
13, 38, 63	HP10136	U937	1409	215
14, 40, 64	HP10175	Stomach cancer	974	112
15, 41, 65	HP10179	KB	925	114
16, 41, 66	HP10196	HT-1080	1115	327
17, 42, 67	HP10235	HT-1080	1721	373
18, 43, 68	HP10297	Stomach cancer	1504	183
19, 44, 69	HP10299	Stomach cancer	532	116
20, 45, 70	HP10301	KB	662	152
21, 46, 71	HP10302	Liver	2373	559
22, 47, 72	HP10304	U-2 OS	1404	330
23, 48, 73	HP10305	U-2 OS	893	108
24, 49, 74	HP10306	U-2 OS	690	101
25, 50, 75	HP10328	KB	2186	372

Hereupon, the same clone as any of the cDNAs of the present invention can be easily obtained by screening of the cDNA library constructed from the cell line or the human tissue employed in the present invention, by the use of an oligonucleotide probe synthesized on the basis of the corresponding cDNA base sequence depicted in Sequence No. 51 to Sequence No. 75.

In general, the polymorphism due to the individual difference is frequently observed in human genes. Therefore, any cDNA that is subjected to insertion or deletion of one or plural nucleotides and/or substitution with other nucleotides in Sequence No. 51 to Sequence No. 75 shall come within the scope of the present invention.

In a similar manner, any protein that is produced by these modifications comprising insertion or deletion of one or plural nucleotides and/or substitution with other nucleotides shall come within the scope of the present invention, as far as said protein possesses the activity of the corresponding protein having the amino acid sequence represented by Sequence No. 1 to Sequence No. 2 or by Sequence No. 4 to Sequence No. 25.

The cDNAs of the present invention include cDNA fragments (more than 10 bp) containing any partial base sequence of the base sequence represented by Sequence No. 26 to No. 50 or of the base sequence represented by Sequence No. 51 to No. 75. Also, DNA fragments consisting of a sense chain and an anti-sense chain shall come within this scope. These DNA fragments can be used as the probes for the gene diagnosis.

#### BRIEF DESCRIPTION OF DRAWINGS

Figure 1: A figure depicting the structure of the secretory signal sequence detection vector pSSD3.

Figure 2: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP00442.



Figure 3: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP00804.

Figure 4: A figure showing the result on the northern-blot hybridization of clone HP00804.

Figure 5: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP01098.

Figure 6: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP01148.

Figure 7: A figure showing the result on the northern-blot hybridization of clone HP01148.

Figure 8: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP01293.

Figure 9: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10013.

Figure 10: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10034.

Figure 11: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10050.

Figure 12: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10071.

Figure 13: A figure depicting the

hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10076.

Figure 14: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10085.

Figure 15: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10122.

Figure 16: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10136.

Figure 17: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10175.

Figure 18: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10179.

Figure 19: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10196.

Figure 20: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10235.

Figure 21: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10297.

Figure 22: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10299.

Figure 23: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10301.

Figure 24: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10302.

Figure 25: A figure depicting the hydrophobicity/hydrophil the protein encoded by clone HP10304.

Figure 26: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10305.

Figure 27: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10306.

Figure 28: A figure depicting the hydrophobicity/hydrophilicity profile of the protein encoded by clone HP10328.

#### BEST MODE FOR CARRING OUT INVENTION

#### EXAMPLE

The present invention is embodied in more detail by the following examples, but this embodiment is not intended to restrict the present invention. The basic operations and the enzyme reactions with regard to the DNA recombination are carried out according to the literature [Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Laboratory, 1989]. Unless otherwise stated, restrictive enzymes and a variety of modification enzymes to be used were those available from

TAKARA SHUZO. The manufacturer's instructions were used for the buffer compositions as well as for the reaction conditions, in each of the enzyme reactions. The cDNA synthesis was carried out according to the literature [Kato, S. et al., Gene 150: 243-250 (1994)].

(1) Preparation of Poly(A)<sup>+</sup> RNA

The fibrosarcoma cell line HT-1080 (ATCC CCL 121), the epidermoid carcinoma cell line KB (ATCC CRL 17), the histiocyte lymphoma cell line U937 (ATCC CRL 1593), the osteosarcoma U-2 OS (ATCC HTB 96), a leukocyte isolated from the peripheral blood, tissues of stomach cancer delivered by the operation, and liver were used for human cells to extract mRNAs. Each of the cell lines was cultured by a conventional procedure.

After about 1 g of human tissues was homogenized in 20 ml of a 5.5 M guanidinium thiocyanate solution, total mRNAs were prepared in accordance with the literature [Okayama, H. et al., "Methods in Enzymology" Vol. 164, Academic Press, 1987]. These mRNAs were subjected to chromatography using an oligo(dT)-cellulose column washed with 20 mM Tris-hydrochloric acid buffer solution (pH 7.6), 0.5 M NaCl, and 1 mM EDTA to obtain a poly(A)<sup>+</sup> RNA in accordance with the above-mentioned literature.

(2) Construction of cDNA Library

To a solution of 10 µg of the above-mentioned poly(A)<sup>+</sup> RNA in 100 mM Tris-hydrochloric acid buffer solution (pH 8) was added one unit of an RNase-free, bacterium-origin alkaline phosphatase and the resulting solution was allowed to react at 37°C for one hour. After the reaction solution

underwent the phenol extraction followed by the ethanol precipitation, the obtained pellets were dissolved in a mixed solution of 50 mM sodium acetate (pH 6), 1 mM EDTA, 0.1% 2-mercaptoethanol, and 0.01% Triton X-100. Thereto was added one unit of a tobacco-origin pyrophosphatase (Epicenter Technologies) and the resulting solution at a total volume of 100  $\mu$ l was allowed to react at 37°C for one hour. After the reaction solution underwent the phenol extraction followed by the ethanol precipitation, the thus-obtained pellets were dissolved in water to obtain a decapped poly(A)<sup>+</sup> RNA solution.

To a solution of the decapped poly(A)<sup>+</sup> RNA and 3 nmol of a DNA-RNA chimeric oligonucleotide (5'-dG-dG-dG-dG-dA-dA-dT-dT-dC-dG-dA-G-G-A-3') in a mixed aqueous solution of 50 mM Tris-hydrochloric acid buffer solution (pH 7.5), 0.5 mM ATP, 5 mM MgCl<sub>2</sub>, 10 mM 2-mercaptoethanol, and 25% polyethylene glycol were added 50 units of T4 RNA ligase and the resulting solution at a total volume of 30  $\mu$ l was allowed to react at 20°C for 12 hours. After the reaction solution underwent the phenol extraction followed by the ethanol precipitation, the thus-obtained pellets were dissolved in water to obtain a chimeric oligo-capped poly(A)<sup>+</sup> RNA.

After the vector pKA1 developed by the present inventors (Japanese Patent Kokai Publication No. 1992-117292) was digested with KpnI, an about 60-dT tail was inserted by a terminal transferase. This product was digested with EcoRV to remove the dT tail at one side and the resulting molecule was used as a vectorial primer.

After 6  $\mu$ g of the previously-prepared chimeric oligo-

capped poly(A)<sup>+</sup> RNA was annealed with 1.2 µg of the vectorial primer, the product was dissolved in a mixed solution of 50 mM Tris-hydrochloric acid buffer solution (pH 8.3), 75 mM KCl, 3 mM MgCl<sub>2</sub>, 10 mM dithiothreitol, and 1.25 mM dNTP (dATP + dCTP + dGTP + dTTP), mixed with 200 units of a reverse transferase (GIBCO-BRL), and the resulting solution at a total volume of 20 µl was allowed to react at 42°C for one hour. After the reaction solution underwent the phenol extraction followed by the ethanol precipitation, the thus-obtained pellets were dissolved in a mixed solution of 50 mM Tris-hydrochloric acid buffer solution (pH 7.5), 100 mM NaCl, 10 mM MgCl<sub>2</sub>, and 1 mM dithiothreitol. Thereto were added 100 units of EcoRI and the resulting solution at a total volume of 20 µl was allowed to react at 37°C for one hour. After the reaction solution underwent the phenol extraction followed by the ethanol precipitation, the obtained pellets were dissolved in a mixed solution of 20 mM Tris-hydrochloric acid buffer solution (pH 7.5), 100 mM KCl, 4 mM MgCl<sub>2</sub>, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, and 50 µg/ml bovine serum albumin. Thereto were added 60 units of *Escherichia coli* DNA ligase and the resulting solution was allowed to react at 16°C for 16 hours. To the reaction solution were added 2 µl of 2 mM dNTP, 4 units of *Escherichia coli* DNA polymerase I, and 0.1 unit of *Escherichia coli* DNase H and the resulting solution was allowed to react at 12°C for one hour and then at 22°C for one hour.

Next, the cDNA-synthesis reaction solution was used to transform *Escherichia coli* DH12S (GIBCO-BRL). The

transformation was carried out by the electroporation method. A portion of the transformant was inoculated on a 2xYT agar culture medium containing 100 µg/ml ampicillin, which was incubated at 37°C overnight. A colony grown on the culture medium was randomly picked up and inoculated on 2 ml of the 2xYT culture medium containing 100 µg/ml ampicillin, which was incubated at 37°C overnight. The culture medium was centrifuged to separate the cells, from which a plasmid DNA was prepared by the alkaline lysis method. After the plasmid DNA was double-digested with EcoRI and NotI, the product was subjected to 0.8% agarose gel electrophoresis to determine the size of the cDNA insert. In addition, by the use of the obtained plasmid as a template, the sequence reaction using M13 universal primer labeled with a fluorescent dye and Taq polymerase (a kit of Applied Biosystems Inc.) was carried out and the product was analyzed by a fluorescent DNA-sequencer (Applied Biosystems Inc.) to determine the base sequence of the cDNA 5'-terminal of about 400 bp. The sequence data were filed as a homo-protein cDNA bank data base.

### (3) Selection of cDNAs Encoding Proteins Having Transmembrane Domains

The base sequence registered in the homo-protein cDNA bank was converted to three frames of amino acid sequences and the presence or absence of an open reading frame (ORF) beginning from the initiation codon. Then, the selection was made for the presence of a signal sequence that is characteristic to a secretory protein at the N-terminal of the portion encoded by ORF. These clones were sequenced from the both 5' and 3' directions by using the deletion method to

determine the whole base sequence. The hydrophobicity/hydrophilicity profiles were obtained for proteins encoded by ORF by the Kyte-Doolittle method [Kyte, J. & Doolittle, R. F., J. Mol. Bio. 157: 105-132 (1982)] to examine the presence or absence of a hydrophobic region. In the case in which there is a hydrophobic region of putative transmembrane domain(s) in the amino acid sequence of an encoded protein, this protein was considered as a membrane protein.

(4) Construction of Secretory Signal Detection Vector pSSD3

One microgram of pSSD1 carrying the SV40 promoter and a cDNA encoding the protease domain of urokinase [Yokoyama-Kobayashi, M. et al., Gene 163: 193-196 (1995)] was digested with 5 units of BglIII and 5 units of EcoRV. Then, after dephosphorylation at the 5' terminal by the CIP treatment, a DNA fragment of about 4.2 kbp was purified by cutting off from the gel of agarose gel electrophoresis.

Two oligo DNA linkers, L1 (5'-GATCCCGGGTCACGTGGGAT-3') and L2 (5'-ATCCACGTGACCCGG-3'), were synthesized and phosphorylated by T4 polynucleotide kinase. After annealing of the both linkers, followed by ligation with the previously-prepared pSSD1 fragment by T4 DNA ligase, *Escherichia coli* JM109 was transformed. A plasmid pSSD3 was prepared from the transformant and the objective recombinant was confirmed by the determination of the base sequence of the linker-inserted fragment. Figure 1 illustrates the structure of the thus-obtained plasmid. The present plasmid vector carries three types of blunt-end formation restriction enzyme sites, SmaI, PmaCI, and EcoRV. Since these cleavage



sites are positioned in succession at an interval of 7 bp, selection of an appropriate site in combination of three types of frames for the inserting cDNA allows to construct a vector expressing a fusion protein.

(5) Functional Verification of Secretory Signal Sequence

Whether the N-terminal hydrophobic region in the secretory protein clone candidate obtained in the above-mentioned steps functions as the secretory signal sequence was verified by the method described in the literature [Yokoyama-Kobayashi, M. et al., Gene 163: 193-196 (1995)]. First, the plasmid containing the target cDNA was cleaved at an appropriate restriction enzyme site that existed at the downstream of the portion expected for encoding the secretory signal sequence. In the case in which this restriction enzyme site was a protruding terminus, the site was blunt-ended by the Klenow treatment or treatment with the mung-bean nuclease. Digestion with HindIII was further carried out and a DNA fragment containing the SV40 promoter and a cDNA encoding the secretory sequence at the downstream of the promoter was separated by agarose gel electrophoresis. This fragment was inserted between the pSSD3 HindIII site and a restriction enzyme site selected so as to match with the urokinase-coding frame, thereby constructing a vector expressing a fusion protein of the secretory signal portion of the target cDNA and the urokinase protease domain.

After *Escherichia coli* (host: JM109) bearing the fusion-protein expression vector was incubated at 37°C for 2 hours in 2 ml of the 2xYT culture medium containing 100 µg/ml ampicillin, the helper phage M13K07 (50 µl) was added and the

incubation was continued at 37°C overnight. A supernatant separated by centrifugation underwent precipitation with polyethylene glycol to obtain single-stranded phage particles. These particles were suspended in 100 µl of 1 mM Tris-0.1 mM EDTA, pH 8 (TE). Also, there was used as a control a suspension of single-stranded particles prepared in the same manner from the vector pKA1-UPA containing pSSD3 and a full-length cDNA of urokinase [Yokoyama-Kobayashi, M. et al., Gene 163: 193-196 (1995)].

The simian-kidney-origin culture cells, COS7, were incubated at 37°C in the presence of 5% CO<sub>2</sub> in the Dulbecco's modified Eagle's culture medium (DMEM) containing 10% fetal calf albumin. Into a 6-well plate (Nunc Inc., 3 cm in the well diameter) were inoculated  $1 \times 10^5$  COS7 cells and incubation was carried out at 37°C for 22 hours in the presence of 5% CO<sub>2</sub>. After the culture medium was removed, the cell surface was washed with a phosphate buffer solution and then washed again with DMEM containing 50 mM Tris-hydrochloric acid (pH 7.5) (TDMEM). To the cells were added 1 µl of the single-stranded phage suspension, 0.6 ml of the DMEM culture medium, and 3 µl of TRANSFECTAM<sup>TM</sup> (IBF Inc.) and the resulting mixture was incubated at 37°C for 3 hours in the presence of 5% CO<sub>2</sub>. After the sample solution was removed, the cell surface was washed with TDMEM, 2 ml per well of DMEM containing 10% fetal calf albumin was added, and the incubation was carried out at 37°C for 2 days in the presence of 5% CO<sub>2</sub>.

To 10 ml of 50 mM phosphate buffer solution (pH 7.4)

containing 2% bovine fibrinogen (Miles Inc.), 0.5% agarose, and 1 mM potassium chloride were added 10 units of human thrombin (Mochida Pharmaceutical Co., Ltd.) and the resulting mixture was solidified in a plate of 9 cm in diameter to prepare a fibrin plate. Ten microliters of the culture supernatant of the transfected COS7 cells were spotted on the fibrin plate, which was incubated at 37°C for 15 hours. The diameter of the thus-obtained clear circle was taken as an index for the urokinase activity. In the case in which a cDNA fragment codes for the amino acid sequence that functions as a secretory signal sequence, a fusion protein is secreted to form a clear circle by its urokinase activity. Therefore, in the case in which a clear circle is not formed, the fusion protein remains as trapped in the membrane and the cDNA fragment is considered to code for a transmembrane domain.

#### (6) Protein Synthesis by In Vitro Translation

The plasmid vector carrying the cDNA of the present invention was utilized for the in vitro transcription/translation by the T<sub>N</sub>T rabbit reticulocyte lysate kit (Promega Biotec). In this case, [<sup>35</sup>S]methionine was added and the expression product was labeled with the radioisotope. All reactions were carried out by following the protocols attached to the kit. Two micrograms of the plasmid was allowed to react at 30°C for 90 minutes in total 25 ml of a reaction solution containing 12.5 µl of the T<sub>N</sub>T rabbit reticulocyte lysate, 0.5 µl of the buffer solution (attached to the kit), 2 µl of an amino acid mixture (methionine-free), 2 µl (0.37 MBq/µl) of [<sup>35</sup>S]methionine (Amersham Corporation), 0.5 µl of T7 RNA polymerase, and 20 U of RNasin. To 3 µl of

the reaction solution was added 2  $\mu$ l of an SDS sampling buffer (125 mM Tris-hydrochloric acid buffer solution, pH 6.8, 120 mM 2-mercaptoethanol, 2% SDS solution, 0.025% bromophenol blue, and 20% glycerol) and the resulting solution was heated at 95°C for 3 minutes and then subjected to SDS-polyacrylamide gel electrophoresis. The molecular weight of the translation product was determined by carrying out the autoradiography.

#### (7) Northern Blot Hybridization

The northern blot hybridization was carried out in order to examine the expression pattern in the human tissues. Membranes on which poly(A)<sup>+</sup> RNAs isolated from each of the human tissues are blotted are purchased from Clontech Inc. cDNA fragments which were excised from the objective clones with appropriate restriction enzymes were subjected to separation by agarose gel electrophoresis followed by labeling with [<sup>32</sup>P] dCPT (Amersham Corporation) using the Random Primer Labeling Kit (Takara Shuzo Co., Ltd.). Hybridization was carried out using a solution attached to the blotted membrane in accordance to the protocol.

#### (8) Expression in COS7

Escherichia coli having an expression vector of the protein of the invention was infected with helper phage M13K07, and single stranded phage was obtained by the above method. Using the thus obtained phage, the expression vector was introduced into simian kidney-originated culture cells COS7 according to the above method. Cultivation was carried out at 37°C in the presence of 5 % CO<sub>2</sub> for 2 hours and then in a medium containing [<sup>35</sup>S]cysteine for 1 hour. The cells

were collected, dissolved and subjected to SDS-PAGE, whereby a band corresponding to a protein as the expression product, which was not present in the COS cells, was revealed.

(9) Clone Examples

<HP00442> (Sequence Number 1, 26, 51)

Determination of the whole base sequence for the cDNA insert of clone HP00442 obtained from the human fibrosarcoma cell line HT-1080 cDNA libraries revealed the structure consisting of a 5'-non-translation region of 81 bp, an ORF of 618 bp, and a 3'-non-translation region of 287 bp. The ORF codes for a protein consisting of 205 amino acid residues with 5 transmembrane domains. Figure 2 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The result of the in vitro translation did not reveal the formation of distinct bands for the translation products and revealed the formation of smeary bands at the high-molecular-weight position.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was analogous to the proteolipid protein PPA1 of the baker's yeast proton ATPase (SWISS-PROT Accession No. P23968). Table 2 indicates the comparison of the amino acid sequences between the human protein of the present invention (HP) and the proteolipid protein PPA1 of the baker's yeast proton ATPase (PL). - represents a gap, \* represents an amino acid residue identical to that in the protein of the present invention, and . represents an amino acid residue analogous to that in the protein of the present invention. The both proteins possessed a homology of 56.8% in the entire region

except for the N-terminal.

Table 2

---

HP	MTGLALLYSGVFVAFWACALAVGVCYTIF-DLGFRFDVAWFLTETSPFMWS
	*..* . * . . . ** ***..*
PL	MNKESKDDDDMSLGKFSFSHFLYVLIVVIVYGLYKLFTGHGSDINFGKFLRTSPYMW
HP	NLGIGLAISLSVVGAAWGIYITGSSIIGGVKAPRIKTKNLVSIIFCEAVAIYGIIMAI
	****.* ..*****.*****.***.***.*****.*****.*****.***
PL	NLGIALCVGLSVVGAAWGIFITGSSMIGAGVRAPRITTKNLISIIIFCEVVAIYGLIIA
HP	ISNMAEPFSATDPKAIGHRNYHAGYSMFGAGLTVGLSNLFCGVCVGIVGSGAALADAQNP
	.*.. **.. . . . * ..***.* **.*** ***.***. ***.*..***.***...
PL	FSSKL--TVATAENMYSKSNLYTGYSLEWAGITVGASNLICGLAVGITGATAAISDAADS
HP	SLFVKILIVEIFGSAIGLFGVIVAILQTSRVKMGD
	..*****.***** .***.*..***.* ...
PL	ALFVKILVIEIFGSILGLLGLIVGLLMAGKASEFQ

---

Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more and also containing the initiation codon (for example, Accession No. H87379), but the present protein can not be predicted from this sequence.

The proteolipid protein PPA1 of the baker's yeast proton ATPase is a membrane protein essential to the growth

of cells [Apperson, M. et al., Biochem. Biophys. Res. Commun. 168: 574-579 (1990)]. Accordingly, the protein of present invention, which is homologous to said protein, is considered to be essential to the growth of human cells and can be utilized for the diagnosis and the treatment of diseases caused by the abnormality of the present protein.

<HP00804> (Sequence Number 2, 27, 52)

Determination of the whole base sequence for the cDNA insert of clone HP00804 obtained from the human leukocyte cell cDNA libraries revealed the structure consisting of a 5'-non-translation region of 132 bp, an ORF of 1116 bp, and a 3'-non-translation region of 576 bp. The ORF codes for a protein consisting of 371 amino acid residues with 7 transmembrane domains. Figure 3 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle. The result of the in vitro translation did not reveal the formation of distinct bands for the translation products.

Examination of the expression pattern in the tissues by the northern blot hybridization using the cDNA fragment of the present invention revealed that the expression occurred in all tissues examined as shown in Figure 4. Therefore, the protein of the present invention is considered to be a housekeeping protein.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was analogous to the rat NMDA receptor - glutamate-binding subunit (GenBank Accession No. S61973). Table 3 indicates the comparison of the amino acid sequences

between the human protein of the present invention (HP) and the rat NMDA receptor - glutamate-binding subunit (RN). - represents a gap, \* represents an amino acid residue identical to that in the protein of the present invention, and represents an amino acid residue analogous to that in the protein of the present invention. This subunit consists of 516 amino acid residues and a region from glutamine at position 68 to arginine at position 342 possessed a 92.6 % homology with the C-terminal 270 amino acid residues in the protein of the present invention. However, any homology was not observed in the N-terminal region. Hereupon, a characteristic repeated sequence that is rich with proline, tyrosine, and glycine was observed in the N-terminal region of the protein of the present invention.

Table 3

---

HP MSHEKSFLVSGDNYPNPGYGGPQPPMPYQAQPPYPGAPYPQPPFQSPYGPQPGYPHG

RN MKRVSWSLGTAILPQTLAILWGHKPLCLPMFSLPTLG

HP PSPYPQGGYPQGPYPQGGYPQGPYPQEGYPQGPYPQGGYPQGPYPQSPFPNYPGQPQVF

\*\*.\*\*\*\*\*. \*

RN PHTHRPLSSPLPMVNQGIPMVPVITRWLPLKDLLKEATHQGHYPQSPFPNYPGQPPPF

HP PGQDPDSPQHGNVQEEGPPSYDNQDFPATNWDDKSIRQAFIRKVFLVLTQLSVTLSTV

\*\*\*.\*\*\*\*\*. \*\* \*\*\*\*\*

RN --QDPGSPQHGNVQEEGPPSYDNQDFPSVNW-DKSIRQAFIRKVFLVLTQLSVTLSTV

HP SVFTFVAEVKGFVRENVWVTVSYAVFFISLIVLSCCGDFRRKHPWNLVALSVLTASLSY



```
..****.*****.*****.*****.*****.*** ****
RN AIFTFVGEVKGFVRANVWTTYVSYAIAFFISLIVLSCCGDFRKKHPWNLVALSILTISLSY
HP MVGMIASFYNTEAVIMAVGITTAVCFTVVIFSMQTRYDFTSCMGVLLVSMVVLFIFAILC
*****.*****
RN MVGMIASFYNTEAVIMAVGITTAVCFTVVIFSMQTRYDFTSCMGVLLVSVVVLFIFAILC
HP IFIRNRILEIVYASLGALLFTCF LAVDTQLLLGKQLSLSP EEVVFAALNLYTDIINIFL
*****
RN IFIRNRILEIVYASLGALLFTCF LAVDTQLLLGKQLSLSP EEVVFAALNLYTDIINIFL
HP YILTIIGRAKE
*****..
RN YILTIIGRSQGIGQAPQVAWVAQTHAPAMTLPSVLPPLWFPAMAWSRGSPSRPRVCTLQ
```

---

Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. W25936), but any of them was shorter than the present cDNA and did not contain the initiation codon.

The rat NMDA receptor - glutamate-binding subunit has been found as one of the subunits of the NMDA receptor complex which exists specifically in the brain [Kumar. K. N. et al., Nature 354: 70-73 (1991)]. Despite a high homology with the protein of the present invention, the subunit shows different expression patterns in the N-terminal sequence and the tissues, whereby both molecules are considered to possess different functions. Since the protein of the present invention possesses 7 transmembrane

domains which are characteristic to channels and transporters, this protein is considered to play a role as a channel and a transporter. Because the protein of the present invention is a housekeeping protein essential to the cells, the present protein can be utilized for the diagnosis and the treatment of diseases caused by the abnormality of this protein.

<HP01098> (Sequence Number 3, 28, 53)

Determination of the whole base sequence for the cDNA insert of clone HP01098 obtained from the human stomach cancer cDNA libraries revealed the structure consisting of a 5'-non-translation region of 61 bp, an ORF of 540 bp, and a 3'-non-translation region of 475 bp. The ORF codes for a protein consisting of 179 amino acid residues with one transmembrane domain. Figure 5 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 20 kDa that was almost consistent with the molecular weight of 20,625 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was completely identical with a 18-kDa subunit of the canine microsomal signal peptidase (SWISS-PROT Accession No. P21378). Therefore, it was verified that the cDNA of the present invention codes for the human homologue of the 18-kDa subunit of the microsomal signal peptidase.

The search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs

possessing the homology of 90% or more (for example, Accession No. T60549), but many sequences were not distinct and the same ORF as that in the present cDNA was not identified.

The 18-kDa subunit of the canine microsomal signal peptidase has been found as one of subunits of the signal peptidase complex that exist in the microsome [Schelness, G. S. & Blobel, G., J. Biol. Chem. 265: 9512-9519 (1990)]. The signal peptidase is an enzyme that cleaves the signal sequence upon secretion of a secretory protein at the endoplasmic reticulum. Therefore, the cDNA of the present invention can be utilized for the production of the present protein as well as for the diagnosis and the treatment of diseases caused by the abnormality of the present protein.  
<HP01148> (Sequence Number 4, 29, 54)

Determination of the whole base sequence for the cDNA insert of clone HP01148 obtained from the human liver cDNA libraries revealed the structure consisting of a 5'-non-translation region of 101 bp, an ORF of 1044 bp, and a 3'-non-translation region of 446 bp. The ORF codes for a protein consisting of 347 amino acid residues with one transmembrane domain at the N-terminal. Figure 6 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. It was indicated that the present protein remained in the membrane from the observation that the urokinase secretion was not identified, upon transduction into the COS7 cells of an expression vector in which a HindIII-PvuII fragment containing a cDNA fragment encoding the N-terminal 178

amino acid residues in the present protein was inserted at the HindIII-PmaCI site of pSSD3. Therefore, the present protein is considered to be a type-II membrane protein. The in vitro translation resulted in the formation of a translation product of 41 kDa that was almost consistent with the molecular weight of 38,101 predicted from the ORF.

Examination of the expression pattern in the tissues by the northern blot hybridization using the cDNA fragment of the present invention revealed that a strong expression occurred in the spleen, as shown in Figure 7. It was also indicated that a slight expression occurred in the liver.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was analogous to the bovine WC1 antigen (SWISS-PROT Accession No. P30205). Table 4 indicates the comparison of the amino acid sequences between the human protein of the present invention (HP) and the bovine WC1 antigen (WC). - represents a gap, \* represents an amino acid residue identical to that in the protein of the present invention, and . represents an amino acid residue analogous to that in the protein of the present invention. The both proteins possessed a homology of 38%.

Table 4

---

HP	MALLFSLILAICTRPGFLASPSGVRVLVGGLHRCCEGRVEVEQKGQWGTVCDDGW
	. . . . . * . * . * . * . * . * . * . * . * . * . * . * . *
WC	VLPQCNDFLSQPAGSAASESSPYCSDSRQLRLVDGGGPGGRVEILDQGSWGTCDDDW

\* .. \* . \* . \*\* . \*\*\*\*\* , \* .....        .        ..     .     . ....\*\*.\*.      .    \*

---

HP YDCSHEEDAGASCENPESFSPVPEGVRLADGPGHCKGRVEVKHQNQWYTV CQTGWSLRA

..\*\*.\*.\*\*\*. \* .\* \* ..\*... .\* \*.\*\* ... \* \*\*. . . .

HP AKVVC RQLGCGRAVLTQKRCNKHAYGRKPTWLSQMSCSGREATLQDCPSGPWGKNTCNHD

..\*..\*\*\*\*\*. .... \*..\* \*.. \*..\* .....\* ..\*\*\*\*\* ..\*..

HP EDTWVECE-----DPFDLRLVGGDNLCSGRLEVLHKGWGSVCDDNWGEKE

\*.....\*\*                      \*\_.\*\*\* \*\*\*, \*\*\*\*\*.\*\* \*.\*\* \*\*,\*\*\*\*\*.\*. \*

HP DOVVCKQLGCGKSLSPSFRDRKCYGPGVGR IWLDNVRCSGEEQSLEQCQHRFWGFHDCTH

..\*\*\*.\*\*\*\*\*..\*..\*..\*\*\*\*.\*\*\*\*\*.\*.\*.\*\*\*.\*\*\* \*\*.\*

HP QEDVAVICSG

\*\*\* \* \*\*\*

WC EEDAGVRCSGVRTTLPTTTAGTRTTSNSLPGIFSLPGVLCILGSLLFLVLVILVTQLLR

The bovine WC1 antigen has been found as a membrane

antigen which is expressed specifically in  $\gamma\delta$  T cells [Wijnngaard, P. L. J. et al., J. Immunol. 149: 3273-3277 (1992)]. The region showing an analogy is called the scavenger receptor cysteine-rich domain (SRCR) which also exists as a repeated sequence in macrophage scavenger receptors [Matsumoto, A. et al., Proc. Natl. Acad. Sci. USA 87: 9133-9137 (1990)], T cell differentiation antigen CD6 [Aruffo, A. et al., J. Exp. Med. 174: 949-952 (1991)], and so on. Since the present protein is expressed specifically in the spleen, This protein is considered to be deeply associated with the functions of the spleen and also to function as a receptor in the same manner as other SRCR family members.

<HP01293> (Sequence Number 5, 30, 55)

Determination of the whole base sequence for the cDNA insert of clone HP01293 obtained from the human liver cDNA libraries revealed the structure consisting of a 5'-non-translation region of 89 bp, an ORF of 1665 bp, and a 3'-non-translation region of 134 bp. The ORF codes for a protein consisting of 554 amino acid residues with 12 transmembrane domains. Figure 8 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation did not reveal the formation of distinct bands and revealed the formation of smeary bands at the high-molecular-weight position.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was analogous to the rat cation transporter

(GenBank Accession No. X78855). Table 5 indicates the comparison of the amino acid sequences between the human protein of the present invention (HP) and the mouse interstitial cell protein (MM). - represents a gap, \* represents an amino acid residue identical to that in the protein of the present invention, and . represents an amino acid residue analogous to that in the protein of the present invention. The both proteins possessed a homology of 78.1% among the entire regions.

Table 5

---

```

HP  MPTVDDILEQVGESGWFKQAFLLCLLSAAFAPICVGIVFLGFTPDHHCQSPGVAELSQ
      *****.***** *****.***.***.***.*****.***.***.*****
RN  MPTVDDVLEQVGEGFWFKQAFLLCLISASLAPIYVGIVFLGFTPGHYCQNPQVAELSQ
HP  RCGWSPAEEELNYTPGLGPAGEA-FLGQCRRYEVDWNQSALSCVDPLASLATNRSHLPLG
      *****.*****.***.***.***.*****.***.***.***.***
RN  RCGWSQAEELNYTPGLGPSDEASFLSQCMRYEVDWNQSTLDCVDPLSSLVANRSQPLPG
HP  PCQDGWVYDTPGSSIVTEFNLVCADSWKLDLFQSCNLACFFFGSLGVGYFADRFRGRKLCL
      **..*****.***.***.***.***.***.***.***.***.***.***
RN  PCEHGWVYDTPGSSIVTEFNLVCGDAWKVDLFQSCVNLGFFLGSLLVGYIADRFRGRKLCL
HP  LGTVLVNAVSGVLMAFSPNYMSMLLFRLLQGLVSKGNWMAGYTLITEFVGSGSRRTVAIM
      * *..***** * ..* *****.*****.***.***.***.***
RN  LVTTLVTSVSGVLTAVAPDYTSMLLFRLLQGMVSKGSWVSGYTLITEFVGSGYRRTTAIL
HP  YQMAFTVGLVALTGLAYALPHWRWLQAVSLPTFLFLYYWCVPEPRWLLSQKRNTETAI
      *****.***.***.***.*****.*****.*****.***.***

```

```

RN  YQMAFTVGLVGLAGVAYAIPDWRWLQAVSLPTFLFLYYWFPESPRWLLSQKRTTRAV
HP  KIMDHIAQKNGKLPADLKMLSLEEDVTEKLSPSFADLFRTPRLRKRTFILMYLWFTDSV
    .**..*****.*****.***..** *****.***.* *****. .*
RN  RIMEQIAQKNGKVPPADLKMLCLEEDASEKRSPSFADLFRTPNLRKHTVILMYLWFSCAV
HP  LYQGLIHEMGATSGNLYLDFLYSALVEIPGAFIALITIDRVGRIYPMVSNLLAGAACL
    *****.*.***..*****.***.***.*.*** *.*****.*****.*.***..*****.
RN  LYQGLIMHVGTGANLYLDFFYSSLVEFPAAFIILVTIDRIGRIYPIAASNLTGAACLL
HP  MIFISPDLEHLNIIIMCVGRMGITIAIQMICLVNAELYPTFVRNLGVMVCSSLCDIGGII
    ****...*****... *.***** **..**.******.*****.*****.***.***.
RN  MIFIPHEHLNVTLACLGRMGATIVLQMVCLVNAELYPTFIRNLGMMVCSALCDLGGIF
HP  TPFIVFRLREVWQALPLILFAVLGLLAAGVTLLLPETKGVALPETMKDAENLG-RKAKPK
    ***.***.*****.***** *...*****.*****.***** **.*.*
RN  TPFMVFRMEVWQALPLILFGVLGLTAGAMTLLLPETKGVALPETIEEAENLGRRKSKAK
HP  ENTIYLVQVTSEPSGT
    *****.*.***...*.
RN  ENTIYLVQVTGKSSST

```

---

Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there did not exist any human gene and human EST possessing the homology of 90% or more.

The rat cation transporter has been found as a membrane protein that relates to the drug excretion in the kidney [Grundemann, D. et al., Nature 372: 549-552 (1994)]. Accordingly, the protein of the present invention which is homologous to this transporter is considered to possess a



similar function and can be utilized for the diagnosis and the treatment of diseases caused by the abnormality of this protein. In addition, since the present protein is considered to relate to the drug excretion, the cells in which this protein is expressed can be utilized as a tool for the drug design of these drugs. Furthermore, since the present protein is expressed principally in the liver and the kidney, a molecule that is prepared so as to possess an affinity to this protein is applicable for the drug delivery system into these tissues.

<HP10013> (Sequence Number 6, 31, 56)

Determination of the whole base sequence for the cDNA insert of clone HP10013 obtained from the human epidermoid carcinoma cell line KB cDNA libraries revealed the structure consisting of a 5'-non-translation region of 96 bp, an ORF of 1053 bp, and a 3'-non-translation region of 884 bp. The ORF codes for a protein consisting of 350 amino acid residues with a signal sequence at the N-terminal and one internal transmembrane domain. Figure 9 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. It was indicated that the present protein functioned as a signal sequence at the N-terminal from the observation that the urokinase activity was detected in the culture medium, upon transduction into the COS7 cells of an expression vector in which a HindIII-Eco065I fragment (treated with the mung-bean nuclease) containing a cDNA fragment encoding the N-terminal 65 amino acid residues in the present protein was inserted at the HindIII-EcoRV site of pSSD3. Therefore, the

present protein is considered to be a type-I membrane protein. The in vitro translation resulted in the formation of a translation product of 39 kDa that was almost consistent with the molecular weight of 39,008 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was not analogous to any of known proteins. Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. H07998), but any of them was shorter than the present cDNA and did not contain the initiation codon.

<HP10034> (Sequence Number 7, 32, 57)

Determination of the whole base sequence for the cDNA insert of clone HP10034 obtained from the human fibrosarcoma cell line HT-1080 cDNA libraries revealed the structure consisting of a 5'-non-translation region of 175 bp, an ORF of 630 bp, and a 3'-non-translation region of 106 bp. The ORF codes for a protein consisting of 209 amino acid residues with 4 transmembrane domains. Figure 10 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 21 kDa that was almost consistent with the molecular weight of 22,432 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was analogous to the human tumor-associated antigen

L6 (SWISS-PROT Accession No. P30408). Table 6 indicates the comparison of the amino acid sequences between the human protein of the present invention (HP) and the human tumor-associated antigen L6 (L6). - represents a gap, \* represents an amino acid residue identical to that in the protein of the present invention, and . represents an amino acid residue analogous to that in the protein of the present invention. The both proteins possessed a homology of 31.8%.

Table 6

---

HP	MVSSPCTQASSRTC	SRILGLSLGTAALFAAGANVALLLPNWDVTYLLRGLLGRHAMLGTG
	*.*	*.* ** . ** . *** . * ** ...* . *.* . . .*
L6	MCYGKCARCIGHSLVGLALLCIAANILLYFPNGETKYASENHL	SRFVWFFSG
HP	LWGGGLMVLTA	-ILISL-MGWRYGCFS--KSGLCRSVLTALLSGGLALLGALICFVTSG
	. ****.*	* ..*.* . ** . ..* ..*....*.. ..* . * ..*
L6	IVGGGLMLLPAFVFIGLEQDDCCGCCGHENCGKRCAMLSSVLAALIGIAGSGYCVIVAA	
HP	VALKDGPFCMF	DVSSFNQTQAWKYGYPFKDLHSRNYLYDRSLWNSVCLEPSAAVVWHVSL
	..* .***.*	* .*. *... .. ** * * *.. * ** . * *.***
L6	LGIAEGPLCL-D-----SLGQWNYTFASTE---	GQYLLDTSTWSE-CTEPKHIVEWNVSL
HP	FSALLCISLLQ	LLLVVHVINSLLGLFCSLCEK
	** ** ..	....* ...***.* ** .*
L6	FSILLALGGIEFILCLIQVINGVLGGICGFCCSHQQQYDC	

---

Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there did not exist any human gene and human EST possessing the homology of 90% or more.

The human tumor-associated antigen L6 is a member of the membrane antigen TM4 super-family proteins that are expressed abundantly on the cell surface of human tumors [Marken, J. S. et al., Proc. Natl. Acad. Sci. USA 89: 3503-3507 (1992)]. Since these membrane antigens are expressed specifically in specific cells and in cancer cells, an antibody that is prepared so as to bind to this antigen is applicable for a variety of diagnoses and as a carrier for the drug delivery. Furthermore, cells in which such a membrane antigen is expressed by transduction of the membrane antigen gene are applicable to the detection of the corresponding ligand.

<HP10050> (Sequence Number 8, 33, 58)

Determination of the whole base sequence for the cDNA insert of clone HP10050 obtained from the human fibrosarcoma cell line HT-1080 cDNA libraries revealed the structure consisting of a 5'-non-translation region of 9 bp, an ORF of 492 bp, and a 3'-non-translation region of 100 bp. The ORF codes for a protein consisting of 163 amino acid residues with one transmembrane domain. Figure 11 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 23 kDa that was almost consistent with the molecular weight of 18,364 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was not analogous to any of known proteins. Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. H03117), but many sequences were not distinct and the same ORF as that in the present cDNA was not identified.

<HP10071> (Sequence Number 9, 34, 59)

Determination of the whole base sequence for the cDNA insert of clone HP10071 obtained from the human stomach cancer cDNA libraries revealed the structure consisting of a 5'-non-translation region of 46 bp, an ORF of 279 bp, and a 3'-non-translation region of 69 bp. The ORF codes for a protein consisting of 92 amino acid residues with 2 transmembrane domains. Figure 12 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 12 kDa that was almost consistent with the molecular weight of 10,094 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was not analogous to any of known proteins. Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. R097442), but many sequences were not

distinct and the same ORF as that in the present cDNA was not identified.

<HP10076> (Sequence Number 10, 35, 60)

Determination of the whole base sequence for the cDNA insert of clone HP10076 obtained from the human lymphoma cell line U937 cDNA libraries revealed the structure consisting of a 5'-non-translation region of 81 bp, an ORF of 519 bp, and a 3'-non-translation region of 132 bp. The ORF codes for a protein consisting of 172 amino acid residues with 2 transmembrane domains. Figure 13 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. It was indicated that the present protein remained in the membrane from the observation that the urokinase secretion was not identified upon transduction into the COS7 cells of an expression vector in which a HindIII-EcoO651 (treated with mung-bean nuclease) fragment containing a cDNA fragment encoding the N-terminal 167 amino acid residues in the present protein was inserted at the HindIII-EcoRV site of pSSD3. The in vitro translation resulted in the formation of a translation product of 24 kDa that was almost consistent with the molecular weight of 18,450 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was analogous to the baker's yeast hypothetical membrane protein of 23.1 kDa (SWISS-PROT Accession No. P34222). Table 7 indicates the comparison of the amino acid sequences between the human protein of the present

invention (HP) and the baker's yeast hypothetical membrane protein of 23.1 kDa (SC). - represents a gap, \* represents an amino acid residue identical to that in the protein of the present invention, and . represents an amino acid residue analogous to that in the protein of the present invention. The both proteins possessed a homology of 47.5% in the C-terminal region of 139 amino acid residues.

Table 7

---

HP	MEYLAHPSTLGLAVGVACGMCLGWS
SC	MITSFLEKMTVSSNYTIALWATFTAISPAVGYQLGTSNASSTKKSSATLLRSKEMKEGK
HP	LRVCFGMLPKSKTSKTHDTDESEASILGD-SGEYKMILVVRNDLKMKGKVAACQSHAAV
	..... *.. *. *.. *.. *.. *.. *.. *.. *.. *.. *..
SC	LHNDTDEEESESEDESEDEDEDIEDIESTSLNDIPGEVRMALVIRQDLGMTKGKIAAQCCHAAL
HP	SAYKQI-----QRRNPEMLKQWEYCGQPKVVVKAPDEETLIALLAHAKMLGLTVSLIQD
	* ...* .. ** * ..* **.*...* **.*. *. *..* **.....*
SC	SCFRHIATNPARASYNPIMTQRWLNAGQAKITLCPDKFTMDELYAKAISLGVNAAVIHD
HP	AGRTQIAPGSQTVLGIQPGPADLIDKVTGHLKLY
	*****.*..*****.*..* ...*..**.*..*
SC	AGRTQIAAGSATVLGLGPAPKAVLDQITGDLKLY

---

Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed

some ESTs possessing the homology of 90% or more (for example, Accession No. T74847), but many sequences were not distinct and the same ORF as that in the present cDNA was not identified.

<HP10085> (Sequence Number 11, 36, 61)

Determination of the whole base sequence for the cDNA insert of clone HP10085 obtained from the human lymphoma cell line U937 cDNA libraries revealed the structure consisting of a 5'-non-translation region of 150 bp, an ORF of 450 bp, and a 3'-non-translation region of 97 bp. The ORF codes for a protein consisting of 149 amino acid residues with one transmembrane domain at the N-terminal. Figure 14 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. It was indicated that the present protein remained in the membrane from the observation that the urokinase secretion was not identified upon transduction into the COS7 cells of an expression vector in which a HindIII-EcoRI fragment (after the Klenow treatment) containing a cDNA fragment encoding the N-terminal 57 amino acid residues in the present protein was inserted at the HindIII-EcoRV site of pSSD3. Therefore, the present protein is considered to be a type-II membrane protein. The in vitro translation resulted in the formation of a translation product of 20 kDa that was almost consistent with the molecular weight of 17,307 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was analogous to the human early activation antigen



CD69 (SWISS-PROT Accession No. Q07108). Table 8 indicates the comparison of the amino acid sequences between the human protein of the present invention (HP) and the human early activation antigen CD69 (CD). - represents a gap, \* represents an amino acid residue identical to that in the protein of the present invention, and . represents an amino acid residue analogous to that in the protein of the present invention. The both proteins possessed a homology of 36.6% in the C-terminal region of 112 amino acid residues.

Table 8

---

HP	MMTKHKKCFI
CD	MSSENCFAENSSLHPESGQENDATSPHFSTRHEGSFQVPVLCVMMNVVFITILIALIA
HP	IVGVLITTNIITLIVKLTRDSQSLCPYDWIGFQNKCYYSKEEGDWNSSKYNCSTQHADL
	* *. **.*.*.*.*.*.*. . *.*.. **.. **
CD	LSVGQYNCPGQYTFSPSDSHVSSCEDWVG YQRKCYFISTVKRSWTS AQNACSEHGATL
HP	TIIDNIEEMNFLRRYKCSSDHWIGLEMAKNRTGQWVDGATFTKSFGMRGSEGCAYLSDDG
	..**.. ..**.*.*. ....*.*.*.*. ....*.*.*.*. ....*.*.*.*. ....
CD	AVIDSEKDMNFLKRYAGREEHWGLKKEPCHPWKWSNGKEFNWPNVTGSDKCVFLKNT
HP	AATARC YTERKWICRKRIH
	... * .. **.*.*
CD	VSSMECEKNLYWICNKP YK

---

Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. H11808), but many sequences are not distinct and the same ORF as that in the present cDNA was not identified.

The human early activation antigen CD69 is a glycoprotein that appears on the surface of activated lymphocytes and a member of the C-type lectin super-family [Hamann, J. et al., J. Immunol. 150: 4920-4927 (1993)]. Since these membrane antigens are expressed specifically in some specific cells, an antibody that is prepared so as to bind to this antigen is applicable for a variety of diagnoses and as a carrier for the drug delivery. Furthermore, cells in which such a membrane antigen is expressed by transduction of the membrane antigen gene are applicable to the detection of the corresponding ligand. <HP10122> (Sequence Number 12, 37, 62)

Determination of the whole base sequence for the cDNA insert of clone HP10122 obtained from the human stomach cancer cDNA libraries revealed the structure consisting of a 5'-non-translation region of 138 bp, an ORF of 567 bp, and a 3'-non-translation region of 481 bp. The ORF codes for a protein consisting of 188 amino acid residues with 2 transmembrane domains. Figure 15 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 22 kDa that was almost consistent with the

molecular weight of 21,175 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was not analogous to any of known proteins. Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. T80360), but many sequences were not distinct and the same ORF as that in the present cDNA was not identified.

<HP10136> (Sequence Number 13, 38, 63)

Determination of the whole base sequence for the cDNA insert of clone HP10136 obtained from the human lymphoma cell line U937 cDNA libraries revealed the structure consisting of a 5'-non-translation region of 81 bp, an ORF of 648 bp, and a 3'-non-translation region of 680 bp. The ORF codes for a protein consisting of 215 amino acid residues with one transmembrane domain at the C-terminal. Figure 16 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 28 kDa that was almost consistent with the molecular weight of 24,740 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was analogous to the baker's yeast protein transport protein SLY2 (SWISS-PROT Accession No. P22214). Table 9 indicates the comparison of the amino acid

sequences between the human protein of the present invention (HP) and the baker's yeast protein transport protein SLY2 (SC). - represents a gap, \* represents an amino acid residue identical to that in the protein of the present invention, and . represents an amino acid residue analogous to that in the protein of the present invention. The both proteins possessed a homology of 36.1% in the entire regions.

Table 9

```

HP  MVLLTMIARVADGLPLAASMQEDEQSGRDLQQYQSQAKQLFRKLNEQSPTRCTLEAGAMT
    *.  *. *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
SC  MIKSTLIYRE-DGLPLCTSVDNENDPS--LFEQKQKVKIVVSRSLTPQSATEATLESGSFE
HP  FHYIEEQGVCYLVLC EAAFPKKLAFAYLEDLHSEFDEQHGKKVPTVS-RPYSFIEFDTFI
    **.  .  *. *. *. *. *. *. *. *. *. *. *. *. *. *. *. *. *. *. *.
SC  IHYLKKSMVYYFVICESGYPRNLAFSYLNDIAQEF EHSFANEYPKPTVRPHYQFVNFDNFL
HP  QKTKKLYIDSRARRNLGSINTELQDVQRI MVANIEEVLQRGEALSALDSKANNLSSL SKK
    *. *. *. *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
SC  QMTKKSYS DKKVQDNLDQLNQLVG VKQIMSKNIEDLLYRGDSL DKMSDMSSSLKETS KR
HP  YRQDAKYLNMRSTYAKLA AVAVFFIMLIVYVRFWWL
    **. *. *. *.  .  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
SC  YRKS AQKINFDLLISQYAPI-VIVAFFVFL-FWWIFLK

```

Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed

some ESTs possessing the homology of 90% or more (for example, Accession No. R80136), but they were shorter than the present cDNA and any molecule containing the initiation codon was not identified.

The baker's yeast protein transport protein SLY2 has been known to be essential for endoplasmic reticulum-to-Golgi protein transport and to be also associated with the control of the cell cycle [Dascher, C. et al., Mol. Cell. Biol. 11: 872-885 (1991)]. Therefore, the cDNA of the present invention can be utilized for the production of the present protein as well as for the diagnosis and the treatment of diseases caused by the abnormality of the present protein.

<HP10175> (Sequence Number 14, 39, 64)

Determination of the whole base sequence for the cDNA insert of clone HP10175 obtained from the human stomach cancer cDNA libraries revealed the structure consisting of a 5'-non-translation region of 173 bp, an ORF of 339 bp, and a 3'-non-translation region of 462 bp. The ORF codes for a protein consisting of 112 amino acid residues with 4 transmembrane domains. Figure 17 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The result of the in vitro translation resulted in the formation of a translation product of 13 kDa that was almost consistent with the molecular weight of 11,564 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was not analogous to any known proteins.

Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. W52852), but many sequences were not distinct and the same ORF as that in the present cDNA was not identified.

<HP10179> (Sequence Number 15, 40, 65)

Determination of the whole base sequence for the cDNA insert of clone HP10179 obtained from the human epidermoid carcinoma cell line KB cDNA libraries revealed the structure consisting of a 5'-non-translation region of 121 bp, an ORF of 345 bp, and a 3'-non-translation region of 459 bp. The ORF codes for a protein consisting of 114 amino acid residues with 4 transmembrane domains. Figure 18 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 14 kDa that was almost consistent with the molecular weight of 12,078 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was not analogous to any known proteins. However, this protein was analogous to the protein encoded by the cDNA clone Hp 10175 of the present invention. Table 10 indicates the comparison of the amino acid sequences between the protein encoded by HP 10179 and the protein encoded by HP 10175. - represents a gap, \* represents an amino acid residue identical to that in the protein of the present invention, and . represents an amino acid residue

analogous to that in the protein of the present invention. The both proteins possessed a homology of 80.8% in the entire regions.

Table 10

---

179 MEKPLFPLVPLHWFGFGYTTALVVSGGIVGYVKTGSVPSLAAGLLFGSLAGLGAYQLYQDP

..\*\*\*\*\*.\*\*\*.\*\*\*.\*\*\*.\*\*\*\*\* \*\*\*

175 MQDTGSVVPLHWFGFGYAALVASGGIIGYVKAGSVPSLAAGLLFGSLAGLGAYQLSQDP

179 RNVWGFLAATSVTFVGVMGMRSYYYGKFMPVGLIAGASLLMAAKVGVRLMTSD

\*\*\*\* \*\* \*\* \*..\*.\*\*\*\* \*. \*\*\*\*\*.\*\*\*\*\*.\*\*\*\*\*.\*.

175 RNVWVFL-ATSGTLAGIMGMRFYHSGKFMPAGLIAGASLLMVAKVGVSMFNRPH

---

Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. N55991), but many sequences were not distinct and the same ORF as that in the present cDNA was not identified.

<HP10196> (Sequence Number 16, 41, 66)

Determination of the whole base sequence for the cDNA insert of clone HP10196 obtained from the human fibrosarcoma cell line HT-1080 cDNA libraries revealed the structure consisting of a 5'-non-translation region of 9 bp, an ORF of 984 bp, and a 3'-non-translation region of 122 bp. The ORF codes for a protein consisting of 327 amino acid residues with one transmembrane domain at the N-

terminal. Figure 19 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. It was indicated that the present protein remained in the membrane from the observation that the urokinase secretion was not identified upon transduction into the COS7 cells of an expression vector in which a HindIII-BglII fragment (after the Klenow treatment) containing a cDNA fragment encoding the N-terminal 162 amino acid residues in the present protein was inserted at the HindIII-EcoRV site of pSSD3. Therefore, the present protein is considered to be a type-II membrane protein. The in vitro translation resulted in the formation of a translation product of 37 kDa that was almost consistent with the molecular weight of 36,163 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was not analogous to any known proteins. Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. T17026), but they were shorter than the present cDNA and any molecule containing the initiation codon was not identified.

<HP10235> (Sequence Number 17, 42, 67)

Determination of the whole base sequence for the cDNA insert of clone HP10235 obtained from the human fibrosarcoma cell line HT-1080 cDNA libraries revealed the structure consisting of a 5'-non-translation region of 5



bp, an ORF of 1122 bp, and a 3'-non-translation region of 594 bp. The ORF codes for a protein consisting of 373 amino acid residues with 11 transmembrane domains. Figure 20 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation did not reveal the formation of distinct bands and revealed the formation of smeary bands at the high-molecular-weight position.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was analogous to the human nucleolar protein HNP36 (EMBL Accession No. X86681). Table 11 indicates the comparison of the amino acid sequences between the human protein of the present invention (HP) and the human nucleolar protein HNP36 (NP). - represents a gap, \* represents an amino acid residue identical to that in the protein of the present invention, and . represents an amino acid residue analogous to that in the protein of the present invention. The both proteins possessed a homology of 45.3% in the entire regions.

Table 11

---

HP	MTLCAMPLLLLFTYLNSFLHQRIPQSVRILGSLVAILLVFLITAILVKVQLDALPFFVIT
HP	MIKIVLINSFGAILQGSFLGLAGLLPASYTAPIMSGQLAGFFASVAMICAIASGSELSE
	* .. .****.*.***** * .*. *.. ..*****.***.***. ....
NP	MASVCFINSFSAVLQGSFLGQLGTMPTSTYSTLFLSGQLAGIFAALAMLLSMASGVDAET



Accordingly, the protein of present invention, which is homologous to said protein, is considered to be a housekeeping protein essential to the growth and multiplication of cells and thereby can be utilized for the diagnosis and the treatment of diseases caused by the abnormality of the present protein.

<HP10297> (Sequence Number 18, 43, 68)

Determination of the whole base sequence for the cDNA insert of clone HP10297 obtained from the human stomach cancer cDNA libraries revealed the structure consisting of a 5'-non-translation region of 62 bp, an ORF of 552 bp, and a 3'-non-translation region of 890 bp. The ORF codes for a protein consisting of 183 amino acid residues with a signal sequence at the N-terminal and one internal transmembrane domain. Therefore, the present protein is considered to be a type-I membrane protein. Figure 21 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 24 kDa that was almost consistent with the molecular weight of 20,574 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was not analogous to any known proteins. Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. R47823), but many sequences are not distinct and the same ORF as that in the present cDNA was not

identified.

<HP10299> (Sequence Number 19, 44, 69)

Determination of the whole base sequence for the cDNA insert of clone HP10299 obtained from the human stomach cancer cDNA libraries revealed the structure consisting of a 5'-non-translation region of 92 bp, an ORF of 351 bp, and a 3'-non-translation region of 89 bp. The ORF codes for a protein consisting of 116 amino acid residues with one transmembrane domain at the N-terminal. Figure 22 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. It was indicated that the present protein remained in the membrane from the observation that the urokinase secretion was not identified upon transduction into the COS7 cells of an expression vector in which a HindIII-VspI fragment (after the Klenow treatment) containing a cDNA fragment encoding the N-terminal 65 amino acid residues in the present protein was inserted at the HindIII-PmaCI site of pSSD3. Therefore, the present protein is considered to be a type-II membrane protein. The in vitro translation resulted in the formation of a translation product of 13 kDa that was almost consistent with the molecular weight of 12,498 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was analogous to the baker's yeast hypothetical membrane protein of 16.5 kDa (SWISS-PROT Accession No. P42834). Table 12 indicates the comparison of the amino acid sequences between the human protein of the present

invention (HP) and the baker's yeast hypothetical membrane protein of 16.5 kDa (SC). - represents a gap, \* represents an amino acid residue identical to that in the protein of the present invention, and . represents an amino acid residue analogous to that in the protein of the present invention. The both proteins possessed a homology of 53.0% in the C-terminal region of 66 amino acid residues.

Table 12

---

HP	MASTVVAVGLTIAAAGFAGRYVLQAMKHMEPVKQVF
SC	MVLPIIIGLGVTMVALSVKSGLNAWTVYKTLSPRTIAKLNNIRIENPTAGYRDALKFKSS
HP	QSLPKSAFSGGYRGGFEPKMTKREAALILGVSP----TANKGKIRDAHRRIMLLNHPDK
	*.****.*.***..** ***...*      .. ... **.*. ****.
SC	LIDEELKNRLNQYQGGFAPRMTEPEALLILDISAREINHLDEKLLKKKHKRKAMVRNHPDR
HP	GGSPYIAAKINEAKDLLEGQAKK
	*****.*****..**
SC	GGSPYMAAKINEAKEVLERSVLLRKR

---

Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. R27748), but many sequences were not distinct and the same ORF as that in the present cDNA was not identified.

<HP10301> (Sequence Number 20, 45, 70)

Determination of the whole base sequence for the cDNA insert of clone HP10301 obtained from the human epidermoid carcinoma cell line KB cDNA libraries revealed the structure consisting of a 5'-non-translation region of 91 bp, an ORF of 459 bp, and a 3'-non-translation region of 112 bp. The ORF codes for a protein consisting of 152 amino acid residues with four transmembrane domains. Figure 23 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 18 kDa that was almost consistent with the molecular weight of 16,516 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was not analogous to any known proteins. Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. N28828), but many sequences were not distinct and the same ORF as that in the present cDNA was not identified.

<HP10302> (Sequence Number 21, 46, 71)

Determination of the whole base sequence for the cDNA insert of clone HP10302 obtained from the human liver cDNA libraries revealed the structure consisting of a 5'-non-translation region of 133 bp, an ORF of 1680 bp, and a 3'-non-translation region of 560 bp. The ORF codes for a protein consisting of 559 amino acid residues with 12

transmembrane domains. Figure 24 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation did not reveal the formation of distinct bands and revealed the formation of smeary bands at the high-molecular-weight position.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was not analogous to any known proteins. Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. N72434), but they were shorter than the present cDNA and any molecule containing the initiation codon was not identified.

<HP10304> (Sequence Number 22, 47, 72)

Determination of the whole base sequence for the cDNA insert of clone HP10304 obtained from the human osteosarcoma U-2 OS cDNA libraries revealed the structure consisting of a 5'-non-translation region of 10 bp, an ORF of 993 bp, and a 3'-non-translation region of 313 bp. The ORF codes for a protein consisting of 330 amino acid residues with a signal sequence at the N-terminal and one internal transmembrane domain. Therefore, the present protein is considered to be a type-I membrane protein. Figure 25 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 36 kDa that was almost

consistent with the molecular weight of 36,840 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was not analogous to any known proteins. Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. N26840), but the same ORF as that in the present cDNA was not identified.

<HP10305> (Sequence Number 23, 48, 73)

Determination of the whole base sequence for the cDNA insert of clone HP10305 obtained from the human osteosarcoma U-2 OS cDNA libraries revealed the structure consisting of a 5'-non-translation region of 109 bp, an ORF of 327 bp, and a 3'-non-translation region of 457 bp. The ORF codes for a protein consisting of 108 amino acid residues with one transmembrane domain. Figure 26 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. It was indicated that the present protein remained in the membrane from the observation that the urokinase secretion was not identified upon transduction into the COS7 cells of an expression vector in which a HindIII-ApaI fragment (treated with mung-bean nuclease) containing a cDNA fragment encoding the N-terminal 162 amino acid residues in the present protein was inserted at the HindIII-PmaCI site of pSSD3. Therefore, the present protein is considered to be a type-II membrane protein. The in vitro translation resulted



in the formation of a translation product of 15 kDa that was almost consistent with the molecular weight of 12,199 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was not analogous to any known proteins. Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. H02768), but many sequences are not distinct and the same ORF as that in the present cDNA was not identified.

<HP10306> (Sequence Number 24, 49, 74)

Determination of the whole base sequence for the cDNA insert of clone HP10306 obtained from the human osteosarcoma U-2 OS cDNA libraries revealed the structure consisting of a 5'-non-translation region of 229 bp, an ORF of 306 bp, and a 3'-non-translation region of 155 bp. The ORF codes for a protein consisting of 101 amino acid residues with 2 transmembrane domains. Figure 27 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. The in vitro translation resulted in the formation of a translation product of 14 kDa that was almost consistent with the molecular weight of 12,029 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the protein was not analogous to any known proteins. Furthermore, the search of GenBank using the base sequence

of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. H44711), but many sequences are not distinct and the same ORF as that in the present cDNA was not identified.

<HP10328> (Sequence Number 25, 50, 75)

Determination of the whole base sequence for the cDNA insert of clone HP10328 obtained from the human epidermoid carcinoma cell line KB cDNA libraries revealed the structure consisting of a 5'-non-translation region of 117 bp, an ORF of 1119 bp, and a 3'-non-translation region of 950 bp. The ORF codes for a protein consisting of 372 amino acid residues with one transmembrane domain. Figure 28 depicts the hydrophobicity/hydrophilicity profile of the present protein obtained by the Kyte-Doolittle method. It was indicated that the present protein remained in the membrane from the observation that the urokinase secretion was not identified upon transduction into the COS7 cells of an expression vector in which a HindIII-PmaCI fragment (treated with mung-bean nuclease) containing a cDNA fragment encoding the N-terminal 129 amino acid residues in the present protein was inserted at the HindIII-SmaI site of pSSD3. Therefore, the present protein is considered to be a type-II membrane protein. The in vitro translation resulted in the formation of a translation product of 41 kDa that was almost consistent with the molecular weight of 42,514 predicted from the ORF.

The search of the protein data base using the amino acid sequence of the present protein revealed that the

protein was analogous to the *Drosophila* neurological secretory signal protein (GenBank Accession No. U41449). Table 13 indicates the comparison of the amino acid sequences between the human protein of the present invention (HP) and the *Drosophila* neurological secretory signal protein (DM). - represents a gap, \* represents an amino acid residue identical to that in the protein of the present invention, and . represents an amino acid residue analogous to that in the protein of the present invention. The both proteins possessed a homology of 38.6% in the middle region of 202 amino acid residues.

Table 13

---

HP	MKYLRHRRPNATLILAIGAFTLLLFSLLVSPPTCKVQEQPPAIPALAWPTPPTRPAPAP	
DM		MQSKHRKLLLRCLLVPLILLVDYCGLLTHL
HP	CHANTSMVTHPDFATQPQHVNFLLYRHCRRHFLLQDVPPSKCAQPVFLLLVIKSSPSNY	
		**. * ..***. *
DM	HELNFERHFFHYPLNDDTGSGSASSGLDKFAYLRVPSFTAEPVVDQPARLTMLIKSAVGNS	
HP	VRRELLRRTWGRERKVRGLQLRLLFLVGTASNPHARKVNRLLEAQTHGDILQWDFHD	
	*** .***** * . ....** .**.***... ..*.. *...***** ** *	
DM	RRREAIRRTWGYEGRFSDVHLRRVFLGTAEDS--EKDVAW----ESREHGDILQADFTD	
HP	SFFNLTLLKQVLFQWQETRCANASFVLNGDDDVFAHTDNMVFYL----QDHDPRHLFVG	
	..** *** . * ..*.... ....* * *** .. ..*.. * *..*.. **.*	
DM	AYFNNTLKTMLGMRWASEQFNRSEFYLFVDDDDYYVSAKNVLKFLGRGRQSHQPE-LLFAG	
HP	QLIQNVGPIRAFWSKYYVPEVVTQNERYPYPYCGGGGFLSRFTAALRRAHVLDIFPID	

...\* ...\* .\*\*\*. . .\*.\*\*\* .\*.\*\*\*. . \* \* .\*.\*\*\*

DM HVFQ-TSPLRHKFSKWYVSLEEYPFDRWPPYVTAGAFILSQKALRQLYAASVHLPLFRFD

HP DVFLGMCLELEGLKPASHSGIRTSGVRAPSQHLSSFDPFCFYRDLLLVRFLPYEMLLMWD

\*\*\*.

DM DVYLGIVALKAGISLQHCDDFRFHRPAYKGPDSYSSVIASHEFGDPEEMTRVWNECRSAN

HP ALNQPNLTCGNQTQIY

DM YA

Furthermore, the search of GenBank using the base sequence of the present cDNA revealed that there existed some ESTs possessing the homology of 90% or more (for example, Accession No. R75815), but they were shorter than the present cDNA and any molecule containing the initiation codon was not identified.

The present invention provides human proteins having transmembrane domains, cDNAs encoding said proteins and eukaryotic cells expressing said cDNA. All of the proteins of the present invention are putative proteins controlling the proliferation and differentiation of the cells, because said proteins exist on the cell membrane. Therefore, the proteins of the present invention can be used as pharmaceuticals or as antigens for preparing antibodies against said proteins. Furthermore, said DNAs can be used for the expression of large amounts of said proteins. The cells expressing large amounts of membrane proteins with transfection of these membrane protein genes can be applied

to the detection of the corresponding ligands, the screening of novel low-molecular medicines, and so on.

In addition to the activities and uses described above, the polynucleotides and proteins of the present invention may exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified below. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or by administration or use of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA).

#### Research Uses and Utilities

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on Southern gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel

polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The proteins provided by the present invention can similarly be used in assay to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the other protein with which binding occurs or to identify inhibitors

of the binding interaction. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E.F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S.L. and A.R. Kimmel eds., 1987.

#### Nutritional Uses

Polynucleotides and proteins of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the protein or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the protein or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

Cytokine and Cell Proliferation/Differentiation  
Activity

A protein of the present invention may exhibit cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+ (preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e and CMK.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., J.



Immunol. 149:3778-3783, 1992; Bowman et al., J. Immunol. 152: 1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A.M. and Shevach, E.M. In Current Protocols in Immunology. J.E.e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human Interferon  $\gamma$ , Schreiber, R.D. In Current Protocols in Immunology. J.E.e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L.S. and Lipsky, P.E. In Current Protocols in Immunology. J.E.e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6 -Nordan, R. In Current Protocols in Immunology. J.E.e.a. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Acad. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11 - Bennett, F., Giannotti, J., Clark, S.C. and Turner, K. J. In Current Protocols in Immunology. J.E.e.a. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and

Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9 - Ciarletta, A., Giannotti, J., Clark, S.C. and Turner, K.J. In Current Protocols in Immunology. J.E.e.a. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immunol. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

#### Immune Stimulating or Suppressing Activity

A protein of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic

activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, *Leishmania* spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, a protein of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein of the present invention may also be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein of the present invention.

Using the proteins of the invention it may also be

possible to immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as , for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration

of a molecule which inhibits or blocks interaction of a B7 lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (e.g., B7-1, B7-3) or blocking antibody), prior to transplantation can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins in vivo as described in Lenschow et al., Science 257:789-792 (1992) and Turka et

al., Proc. Natl. Acad. Sci USA, 89:11102-11105 (1992). In addition, murine models of GVHD (see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function in vivo on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor:ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which may be involved in the disease process. Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine experimental autoimmune encephalitis, systemic lupus erythmatosis in MRL/lpr/lpr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in NOD mice and BB rats, and murine experimental myasthenia gravis

(see Paul ed., Fundamental Immunology, Raven Press, New York, 1989, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of immune responses may be in the form of enhancing an existing immune response or eliciting an initial immune response. For example, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory forms of B lymphocyte antigens systemically.

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

In another application, up regulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells (e.g., sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected ex vivo with an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection in vivo.

The presence of the peptide of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I  $\alpha$  chain protein and  $\beta_2$  microglobulin protein or an



MHC class II $\alpha$  chain protein and an MHC class II $\beta$  chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J.

Immunol. 140:508-512, 1988; Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., J. Immunol. 137:3494-3500, 1986; Bowman et al., J. Virology 61:1992-1998; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro antibody production, Mond, J.J. and Brunswick, M. In Current Protocols in Immunology. J.E.e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify,

among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995;

Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

Hematopoiesis Regulating Activity

A protein of the present invention may be useful in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell deficiencies. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without

limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either in-vivo or ex-vivo (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hematopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M.G. In Culture of Hematopoietic Cells. R.I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, NY. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I.K. and Briddell, R.A. In Culture of Hematopoietic Cells. R.I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss,

Inc., New York, NY. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R.E. In Culture of Hematopoietic Cells. R.I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, NY. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R.I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, NY. 1994; Long term culture initiating cell assay, Sutherland, H.J. In Culture of Hematopoietic Cells. R.I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, NY. 1994.

#### Tissue Growth Activity

A protein of the present invention also may have utility in compositions used for bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as for wound healing and tissue repair and replacement, and in the treatment of burns, incisions and ulcers.

A protein of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced

craniofacial defects, and also is useful in cosmetic plastic surgery.

A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of progenitors of bone-forming cells. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein of the present invention is tendon/ligament formation. A protein of the present invention, which induces tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or

other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors ex vivo for return in vivo to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic



disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to regenerate. A protein of the invention may also exhibit angiogenic activity.

A protein of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium ).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, HI and Rovee, DT, eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

#### Activin/Inhibin Activity

A protein of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a protein of the present invention, alone or in heterodimers with a member of the inhibin  $\alpha$  family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin- $\beta$  group, may be useful as a fertility inducing

therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, United States Patent 4,798,885. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., *Endocrinology* 91:562-572, 1972; Ling et al., *Nature* 321:779-782, 1986; Vale et al., *Nature* 321:776-779, 1986; Mason et al., *Nature* 318:659-663, 1985; Forage et al., *Proc. Natl. Acad. Sci. USA* 83:3091-3095, 1986.

#### Chemotactic/Chemokinetic Activity

A protein of the present invention may have chemotactic or chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells.

Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of

infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J.E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25: 1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153: 1762-1768, 1994.

### Hemostatic and Thrombolytic Activity

A protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke)).

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

### Receptor/Ligand Activity

A protein of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular

adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J.E. Coligan, A.M. Kruisbeek, D.H. Margulies, E.M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1-7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

#### Anti-Inflammatory Activity

Proteins of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting

cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation inflammation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Proteins of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material.

#### Tumor Inhibition Activity

In addition to the activities described above for immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other

factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or inhibiting factors, agents or cell types which promote tumor growth

Other Activities

A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating



deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

## SEQUENCE LISTING

Sequence No.: 1

Sequence length: 205

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Fibrosarcoma

Cell line: HT-1080

Clone name: HP00442

Sequence description

```

Met Thr Gly Leu Ala Leu Leu Tyr Ser Gly Val Phe Val Ala Phe Trp
  1             5             10             15
Ala Cys Ala Leu Ala Val Gly Val Cys Tyr Thr Ile Phe Asp Leu Gly
      20             25             30
Phe Arg Phe Asp Val Ala Trp Phe Leu Thr Glu Thr Ser Pro Phe Met
      35             40             45
Trp Ser Asn Leu Gly Ile Gly Leu Ala Ile Ser Leu Ser Val Val Gly
      50             55             60
Ala Ala Trp Gly Ile Tyr Ile Thr Gly Ser Ser Ile Ile Gly Gly Gly
      65             70             75             80
Val Lys Ala Pro Arg Ile Lys Thr Lys Asn Leu Val Ser Ile Ile Phe
      85             90             95
Cys Glu Ala Val Ala Ile Tyr Gly Ile Ile Met Ala Ile Val Ile Ser
      100            105            110
Asn Met Ala Glu Pro Phe Ser Ala Thr Asp Pro Lys Ala Ile Gly His
      115            120            125
Arg Asn Tyr His Ala Gly Tyr Ser Met Phe Gly Ala Gly Leu Thr Val
      130            135            140
Gly Leu Ser Asn Leu Phe Cys Gly Val Cys Val Gly Ile Val Gly Ser
      145            150            155            160
Gly Ala Ala Leu Ala Asp Ala Gln Asn Pro Ser Leu Phe Val Lys Ile
      165            170            175
Leu Ile Val Glu Ile Phe Gly Ser Ala Ile Gly Leu Phe Gly Val Ile
      180            185            190
Val Ala Ile Leu Gln Thr Ser Arg Val Lys Met Gly Asp
      195            200            205

```

Sequence No.: 2

Sequence length: 371

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Leukocyte

Clone name: HP00804

Sequence description

```

Met Ser His Glu Lys Ser Phe Leu Val Ser Gly Asp Asn Tyr Pro Pro
 1             5             10             15
Pro Asn Pro Gly Tyr Pro Gly Gly Pro Gln Pro Pro Met Pro Pro Tyr
      20             25             30
Ala Gln Pro Pro Tyr Pro Gly Ala Pro Tyr Pro Gln Pro Pro Phe Gln
      35             40             45
Pro Ser Pro Tyr Gly Gln Pro Gly Tyr Pro His Gly Pro Ser Pro Tyr
      50             55             60
Pro Gln Gly Gly Tyr Pro Gln Gly Pro Tyr Pro Gln Gly Gly Tyr Pro
      65             70             75             80
Gln Gly Pro Tyr Pro Gln Glu Gly Tyr Pro Gln Gly Pro Tyr Pro Gln
      85             90             95
Gly Gly Tyr Pro Gln Gly Pro Tyr Pro Gln Ser Pro Phe Pro Pro Asn
      100             105             110
Pro Tyr Gly Gln Pro Gln Val Phe Pro Gly Gln Asp Pro Asp Ser Pro
      115             120             125
Gln His Gly Asn Tyr Gln Glu Glu Gly Pro Pro Ser Tyr Tyr Asp Asn
      130             135             140
Gln Asp Phe Pro Ala Thr Asn Trp Asp Asp Lys Ser Ile Arg Gln Ala
      145             150             155             160
Phe Ile Arg Lys Val Phe Leu Val Leu Thr Leu Gln Leu Ser Val Thr
      165             170             175
Leu Ser Thr Val Ser Val Phe Thr Phe Val Ala Glu Val Lys Gly Phe
      180             185             190
Val Arg Glu Asn Val Trp Thr Tyr Tyr Val Ser Tyr Ala Val Phe Phe
      195             200             205
Ile Ser Leu Ile Val Leu Ser Cys Cys Gly Asp Phe Arg Arg Lys His
      210             215             220
Pro Trp Asn Leu Val Ala Leu Ser Val Leu Thr Ala Ser Leu Ser Tyr
      225             230             235             240
Met Val Gly Met Ile Ala Ser Phe Tyr Asn Thr Glu Ala Val Ile Met
      245             250             255

```

```

Ala Val Gly Ile Thr Thr Ala Val Cys Phe Thr Val Val Ile Phe Ser
      260                      265                      270
Met Gln Thr Arg Tyr Asp Phe Thr Ser Cys Met Gly Val Leu Leu Val
      275                      280                      285
Ser Met Val Val Leu Phe Ile Phe Ala Ile Leu Cys Ile Phe Ile Arg
      290                      295                      300
Asn Arg Ile Leu Glu Ile Val Tyr Ala Ser Leu Gly Ala Leu Leu Phe
      305                      310                      315                      320
Thr Cys Phe Leu Ala Val Asp Thr Gln Leu Leu Leu Gly Asn Lys Gln
      325                      330                      335
Leu Ser Leu Ser Pro Glu Glu Tyr Val Phe Ala Ala Leu Asn Leu Tyr
      340                      345                      350
Thr Asp Ile Ile Asn Ile Phe Leu Tyr Ile Leu Thr Ile Ile Gly Arg
      355                      360                      365
Ala Lys Glu
      370

```

Sequence No.: 3

Sequence length: 179

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP01098

Sequence description

```

Met Leu Ser Leu Asp Phe Leu Asp Asp Val Arg Arg Met Asn Lys Arg
  1                      5                      10                      15
Gln Leu Tyr Tyr Gln Val Leu Asn Phe Gly Met Ile Val Ser Ser Ala
      20                      25                      30
Leu Met Ile Trp Lys Gly Leu Met Val Ile Thr Gly Ser Glu Ser Pro
      35                      40                      45
Ile Val Val Val Leu Ser Gly Ser Met Glu Pro Ala Phe His Arg Gly
      50                      55                      60
Asp Leu Leu Phe Leu Thr Asn Arg Val Glu Asp Pro Ile Arg Val Gly
      65                      70                      75                      80
Glu Ile Val Val Phe Arg Ile Glu Gly Arg Glu Ile Pro Ile Val His
      85                      90                      95
Arg Val Leu Lys Ile His Glu Lys Gln Asn Gly His Ile Lys Phe Leu
      100                      105                      110

```

Thr Lys Gly Asp Asn Asn Ala Val Asp Asp Arg Gly Leu Tyr Lys Gln  
 115 120 125  
 Gly Gln His Trp Leu Glu Lys Lys Asp Val Val Gly Arg Ala Arg Gly  
 130 135 140  
 Phe Val Pro Tyr Ile Gly Ile Val Thr Ile Leu Met Asn Asp Tyr Pro  
 145 150 155 160  
 Lys Phe Lys Tyr Ala Val Leu Phe Leu Leu Gly Leu Phe Val Leu Val  
 165 170 175  
 His Arg Glu

Sequence No.: 4

Sequence length: 347

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Liver

Clone name: HP01148

Sequence description

Met Ala Leu Leu Phe Ser Leu Ile Leu Ala Ile Cys Thr Arg Pro Gly  
 1 5 10 15  
 Phe Leu Ala Ser Pro Ser Gly Val Arg Leu Val Gly Gly Leu His Arg  
 20 25 30  
 Cys Glu Gly Arg Val Glu Val Glu Gln Lys Gly Gln Trp Gly Thr Val  
 35 40 45  
 Cys Asp Asp Gly Trp Asp Ile Lys Asp Val Ala Val Leu Cys Arg Glu  
 50 55 60  
 Leu Gly Cys Gly Ala Ala Ser Gly Thr Pro Ser Gly Ile Leu Tyr Glu  
 65 70 75 80  
 Pro Pro Ala Glu Lys Glu Gln Lys Val Leu Ile Gln Ser Val Ser Cys  
 85 90 95  
 Thr Gly Thr Glu Asp Thr Leu Ala Gln Cys Glu Gln Glu Glu Val Tyr  
 100 105 110  
 Asp Cys Ser His Glu Glu Asp Ala Gly Ala Ser Cys Glu Asn Pro Glu  
 115 120 125  
 Ser Ser Phe Ser Pro Val Pro Glu Gly Val Arg Leu Ala Asp Gly Pro  
 130 135 140  
 Gly His Cys Lys Gly Arg Val Glu Val Lys His Gln Asn Gln Trp Tyr  
 145 150 155 160  
 Thr Val Cys Gln Thr Gly Trp Ser Leu Arg Ala Ala Lys Val Val Cys

96

	165		170		175
Arg Gln Leu Gly Cys Gly Arg Ala Val Leu Thr Gln Lys Arg Cys Asn					
	180		185		190
Lys His Ala Tyr Gly Arg Lys Pro Ile Trp Leu Ser Gln Met Ser Cys					
	195		200		205
Ser Gly Arg Glu Ala Thr Leu Gln Asp Cys Pro Ser Gly Pro Trp Gly					
	210		215		220
Lys Asn Thr Cys Asn His Asp Glu Asp Thr Trp Val Glu Cys Glu Asp					
225		230		235	240
Pro Phe Asp Leu Arg Leu Val Gly Gly Asp Asn Leu Cys Ser Gly Arg					
	245		250		255
Leu Glu Val Leu His Lys Gly Val Trp Gly Ser Val Cys Asp Asp Asn					
	260		265		270
Trp Gly Glu Lys Glu Asp Gln Val Val Cys Lys Gln Leu Gly Cys Gly					
	275		280		285
Lys Ser Leu Ser Pro Ser Phe Arg Asp Arg Lys Cys Tyr Gly Pro Gly					
	290		295		300
Val Gly Arg Ile Trp Leu Asp Asn Val Arg Cys Ser Gly Glu Glu Gln					
305		310		315	320
Ser Leu Glu Gln Cys Gln His Arg Phe Trp Gly Phe His Asp Cys Thr					
	325		330		335
His Gln Glu Asp Val Ala Val Ile Cys Ser Gly					
	340		345		

Sequence No.: 5

Sequence length: 554

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Liver

Clone name: HP01293

Sequence description

Met Pro Thr Val Asp Asp Ile Leu Glu Gln Val Gly Glu Ser Gly Trp			
1	5	10	15
Phe Gln Lys Gln Ala Phe Leu Ile Leu Cys Leu Leu Ser Ala Ala Phe			
20	25	30	
Ala Pro Ile Cys Val Gly Ile Val Phe Leu Gly Phe Thr Pro Asp His			
35	40	45	
His Cys Gln Ser Pro Gly Val Ala Glu Leu Ser Gln Arg Cys Gly Trp			

50	55	60
Ser Pro Ala Glu Glu Leu Asn Tyr Thr Val Pro Gly Leu Gly Pro Ala		
65	70	75
Gly Glu Ala Phe Leu Gly Gln Cys Arg Arg Tyr Glu Val Asp Trp Asn		
	85	90
Gln Ser Ala Leu Ser Cys Val Asp Pro Leu Ala Ser Leu Ala Thr Asn		
	100	105
Arg Ser His Leu Pro Leu Gly Pro Cys Gln Asp Gly Trp Val Tyr Asp		
	115	120
Thr Pro Gly Ser Ser Ile Val Thr Glu Phe Asn Leu Val Cys Ala Asp		
	130	135
Ser Trp Lys Leu Asp Leu Phe Gln Ser Cys Leu Asn Ala Gly Phe Phe		
	145	150
Phe Gly Ser Leu Gly Val Gly Tyr Phe Ala Asp Arg Phe Gly Arg Lys		
	165	170
Leu Cys Leu Leu Gly Thr Val Leu Val Asn Ala Val Ser Gly Val Leu		
	180	185
Met Ala Phe Ser Pro Asn Tyr Met Ser Met Leu Leu Phe Arg Leu Leu		
	195	200
Gln Gly Leu Val Ser Lys Gly Asn Trp Met Ala Gly Tyr Thr Leu Ile		
	210	215
Thr Glu Phe Val Gly Ser Gly Ser Arg Arg Thr Val Ala Ile Met Tyr		
	225	230
Gln Met Ala Phe Thr Val Gly Leu Val Ala Leu Thr Gly Leu Ala Tyr		
	245	250
Ala Leu Pro His Trp Arg Trp Leu Gln Leu Ala Val Ser Leu Pro Thr		
	260	265
Phe Leu Phe Leu Leu Tyr Tyr Trp Cys Val Pro Glu Ser Pro Arg Trp		
	275	280
Leu Leu Ser Gln Lys Arg Asn Thr Glu Ala Ile Lys Ile Met Asp His		
	290	295
Ile Ala Gln Lys Asn Gly Lys Leu Pro Pro Ala Asp Leu Lys Met Leu		
	305	310
Ser Leu Glu Glu Asp Val Thr Glu Lys Leu Ser Pro Ser Phe Ala Asp		
	325	330
Leu Phe Arg Thr Pro Arg Leu Arg Lys Arg Thr Phe Ile Leu Met Tyr		
	340	345
Leu Trp Phe Thr Asp Ser Val Leu Tyr Gln Gly Leu Ile Leu His Met		
	355	360
Gly Ala Thr Ser Gly Asn Leu Tyr Leu Asp Phe Leu Tyr Ser Ala Leu		
	370	375
Val Glu Ile Pro Gly Ala Phe Ile Ala Leu Ile Thr Ile Asp Arg Val		
	385	390
Gly Arg Ile Tyr Pro Met Ala Val Ser Asn Leu Leu Ala Gly Ala Ala		
	395	400

	405		410		415										
Cys	Leu	Val	Met	Ile	Phe	Ile	Ser	Pro	Asp	Leu	His	Trp	Leu	Asn	Ile
	420		425		430										
Ile	Ile	Met	Cys	Val	Gly	Arg	Met	Gly	Ile	Thr	Ile	Ala	Ile	Gln	Met
	435		440		445										
Ile	Cys	Leu	Val	Asn	Ala	Glu	Leu	Tyr	Pro	Thr	Phe	Val	Arg	Asn	Leu
	450		455		460										
Gly	Val	Met	Val	Cys	Ser	Ser	Leu	Cys	Asp	Ile	Gly	Gly	Ile	Ile	Thr
465			470		475									480	
Pro	Phe	Ile	Val	Phe	Arg	Leu	Arg	Glu	Val	Trp	Gln	Ala	Leu	Pro	Leu
	485		490		495										
Ile	Leu	Phe	Ala	Val	Leu	Gly	Leu	Leu	Ala	Ala	Gly	Val	Thr	Leu	Leu
	500		505		510										
Leu	Pro	Glu	Thr	Lys	Gly	Val	Ala	Leu	Pro	Glu	Thr	Met	Lys	Asp	Ala
	515		520		525										
Glu	Asn	Leu	Gly	Arg	Lys	Ala	Lys	Pro	Lys	Glu	Asn	Thr	Ile	Tyr	Leu
	530		535		540										
Lys	Val	Gln	Thr	Ser	Glu	Pro	Ser	Gly	Thr						
545			550												

Sequence No.: 6

Sequence length: 350

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP10013

Sequence description

Met	Ala	Val	Phe	Val	Val	Leu	Leu	Ala	Leu	Val	Ala	Gly	Val	Leu	Gly
1			5						10					15	
Asn	Glu	Phe	Ser	Ile	Leu	Lys	Ser	Pro	Gly	Ser	Val	Val	Phe	Arg	Asn
	20							25					30		
Gly	Asn	Trp	Pro	Ile	Pro	Gly	Glu	Arg	Ile	Pro	Asp	Val	Ala	Ala	Leu
	35					40					45				
Ser	Met	Gly	Phe	Ser	Val	Lys	Glu	Asp	Leu	Ser	Trp	Pro	Gly	Leu	Ala
	50					55					60				
Val	Gly	Asn	Leu	Phe	His	Arg	Pro	Arg	Ala	Thr	Val	Met	Val	Met	Val
65			70						75					80	



Lys Gly Val Asn Lys Leu Ala Leu Pro Pro Gly Ser Val Ile Ser Tyr			
	85	90	95
Pro Leu Glu Asn Ala Val Pro Phe Ser Leu Asp Ser Val Ala Asn Ser			
	100	105	110
Ile His Ser Leu Phe Ser Glu Glu Thr Pro Val Val Leu Gln Leu Ala			
	115	120	125
Pro Ser Glu Glu Arg Val Tyr Met Val Gly Lys Ala Asn Ser Val Phe			
	130	135	140
Glu Asp Leu Ser Val Thr Leu Arg Gln Leu Arg Asn Arg Leu Phe Gln			
	145	150	155
Glu Asn Ser Val Leu Ser Ser Leu Pro Leu Asn Ser Leu Ser Arg Asn			
	165	170	175
Asn Glu Val Asp Leu Leu Phe Leu Ser Glu Leu Gln Val Leu His Asp			
	180	185	190
Ile Ser Ser Leu Leu Ser Arg His Lys His Leu Ala Lys Asp His Ser			
	195	200	205
Pro Asp Leu Tyr Ser Leu Glu Leu Ala Gly Leu Asp Glu Ile Gly Lys			
	210	215	220
Arg Tyr Gly Glu Asp Ser Glu Gln Phe Arg Asp Ala Ser Lys Ile Leu			
	225	230	235
Val Asp Ala Leu Gln Lys Phe Ala Asp Asp Met Tyr Ser Leu Tyr Gly			
	245	250	255
Gly Asn Ala Val Val Glu Leu Val Thr Val Lys Ser Phe Asp Thr Ser			
	260	265	270
Leu Ile Arg Lys Thr Arg Thr Ile Leu Glu Ala Lys Gln Ala Lys Asn			
	275	280	285
Pro Ala Ser Pro Tyr Asn Leu Ala Tyr Lys Tyr Asn Phe Glu Tyr Ser			
	290	295	300
Val Val Phe Asn Met Val Leu Trp Ile Met Ile Ala Leu Ala Leu Ala			
	305	310	315
Val Ile Ile Thr Ser Tyr Asn Ile Trp Asn Met Asp Pro Gly Tyr Asp			
	325	330	335
Ser Ile Ile Tyr Arg Met Thr Asn Gln Lys Ile Arg Met Asp			
	340	345	350

Sequence No.: 7

Sequence length: 209

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Fibrosarcoma

Cell line: HT-1080

Clone name: HP10034

## Sequence description

```

Met Val Ser Ser Pro Cys Thr Gln Ala Ser Ser Arg Thr Cys Ser Arg
 1             5             10             15
Ile Leu Gly Leu Ser Leu Gly Thr Ala Ala Leu Phe Ala Ala Gly Ala
      20             25             30
Asn Val Ala Leu Leu Leu Pro Asn Trp Asp Val Thr Tyr Leu Leu Arg
      35             40             45
Gly Leu Leu Gly Arg His Ala Met Leu Gly Thr Gly Leu Trp Gly Gly
 50             55             60
Gly Leu Met Val Leu Thr Ala Ala Ile Leu Ile Ser Leu Met Gly Trp
 65             70             75             80
Arg Tyr Gly Cys Phe Ser Lys Ser Gly Leu Cys Arg Ser Val Leu Thr
      85             90             95
Ala Leu Leu Ser Gly Gly Leu Ala Leu Leu Gly Ala Leu Ile Cys Phe
      100             105             110
Val Thr Ser Gly Val Ala Leu Lys Asp Gly Pro Phe Cys Met Phe Asp
      115             120             125
Val Ser Ser Phe Asn Gln Thr Gln Ala Trp Lys Tyr Gly Tyr Pro Phe
      130             135             140
Lys Asp Leu His Ser Arg Asn Tyr Leu Tyr Asp Arg Ser Leu Trp Asn
      145             150             155             160
Ser Val Cys Leu Glu Pro Ser Ala Ala Val Val Trp His Val Ser Leu
      165             170             175
Phe Ser Ala Leu Leu Cys Ile Ser Leu Leu Gln Leu Leu Leu Val Val
      180             185             190
Val His Val Ile Asn Ser Leu Leu Gly Leu Phe Cys Ser Leu Cys Glu
      195             200             205
Lys

```

Sequence No.: 8

Sequence length: 163

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Fibrosarcoma

Cell line: HT-1080

Clone name: HP10050

## Sequence description

```

Met Ala Ala Gly Leu Phe Gly Leu Ser Ala Arg Arg Leu Leu Ala Ala
 1             5             10             15
Ala Ala Thr Arg Gly Leu Pro Ala Ala Arg Val Arg Trp Glu Ser Ser
          20             25             30
Phe Ser Arg Thr Val Val Ala Pro Ser Ala Val Ala Gly Lys Arg Pro
      35             40             45
Pro Glu Pro Thr Thr Pro Trp Gln Glu Asp Pro Glu Pro Glu Asp Glu
      50             55             60
Asn Leu Tyr Glu Lys Asn Pro Asp Ser His Gly Tyr Asp Lys Asp Pro
      65             70             75             80
Val Leu Asp Val Trp Asn Met Arg Leu Val Phe Phe Phe Gly Val Ser
          85             90             95
Ile Ile Leu Val Leu Gly Ser Thr Phe Val Ala Tyr Leu Pro Asp Tyr
      100             105             110
Arg Cys Thr Gly Cys Pro Arg Ala Trp Asp Gly Met Lys Glu Trp Ser
      115             120             125
Arg Arg Glu Ala Glu Arg Leu Val Lys Tyr Arg Glu Ala Asn Gly Leu
      130             135             140
Pro Ile Met Glu Ser Asn Cys Phe Asp Pro Ser Lys Ile Gln Leu Pro
      145             150             155             160
Glu Asp Glu

```

Sequence No.: 9

Sequence length: 92

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10071

## Sequence description

```

Met Thr Lys Leu Ala Gln Trp Leu Trp Gly Leu Ala Ile Leu Gly Ser
 1             5             10             15
Thr Trp Val Ala Leu Thr Thr Gly Ala Leu Gly Leu Glu Leu Pro Leu
          20             25             30
Ser Cys Gln Glu Val Leu Trp Pro Leu Pro Ala Tyr Leu Leu Val Ser
      35             40             45

```

Ala Gly Cys Tyr Ala Leu Gly Thr Val Gly Tyr Arg Val Ala Thr Phe  
 50 55 60  
 His Asp Cys Glu Asp Ala Ala Arg Glu Leu Gln Ser Gln Ile Gln Glu  
 65 70 75 80  
 Ala Arg Ala Asp Leu Ala Arg Arg Gly Leu Arg Phe  
 85 90

Sequence No.: 10

Sequence length: 172

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Lymphoma

Cell line: U937

Clone name: HP10076

Sequence description

Met Glu Tyr Leu Ala His Pro Ser Thr Leu Gly Leu Ala Val Gly Val  
 1 5 10 15  
 Ala Cys Gly Met Cys Leu Gly Trp Ser Leu Arg Val Cys Phe Gly Met  
 20 25 30  
 Leu Pro Lys Ser Lys Thr Ser Lys Thr His Thr Asp Thr Glu Ser Glu  
 35 40 45  
 Ala Ser Ile Leu Gly Asp Ser Gly Glu Tyr Lys Met Ile Leu Val Val  
 50 55 60  
 Arg Asn Asp Leu Lys Met Gly Lys Gly Lys Val Ala Ala Gln Cys Ser  
 65 70 75 80  
 His Ala Ala Val Ser Ala Tyr Lys Gln Ile Gln Arg Arg Asn Pro Glu  
 85 90 95  
 Met Leu Lys Gln Trp Glu Tyr Cys Gly Gln Pro Lys Val Val Val Lys  
 100 105 110  
 Ala Pro Asp Glu Glu Thr Leu Ile Ala Leu Leu Ala His Ala Lys Met  
 115 120 125  
 Leu Gly Leu Thr Val Ser Leu Ile Gln Asp Ala Gly Arg Thr Gln Ile  
 130 135 140  
 Ala Pro Gly Ser Gln Thr Val Leu Gly Ile Gly Pro Gly Pro Ala Asp  
 145 150 155 160  
 Leu Ile Asp Lys Val Thr Gly His Leu Lys Leu Tyr  
 165 170

Sequence No.: 11

Sequence length: 149

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Lymphoma

Cell line: U937

Clone name: HP10085

Sequence description

```

Met Met Thr Lys His Lys Lys Cys Phe Ile Ile Val Gly Val Leu Ile
 1             5             10             15
Thr Thr Asn Ile Ile Thr Leu Ile Val Lys Leu Thr Arg Asp Ser Gln
          20             25             30
Ser Leu Cys Pro Tyr Asp Trp Ile Gly Phe Gln Asn Lys Cys Tyr Tyr
          35             40             45
Phe Ser Lys Glu Glu Gly Asp Trp Asn Ser Ser Lys Tyr Asn Cys Ser
          50             55             60
Thr Gln His Ala Asp Leu Thr Ile Ile Asp Asn Ile Glu Glu Met Asn
 65             70             75             80
Phe Leu Arg Arg Tyr Lys Cys Ser Ser Asp His Trp Ile Gly Leu Lys
          85             90             95
Met Ala Lys Asn Arg Thr Gly Gln Trp Val Asp Gly Ala Thr Phe Thr
          100            105            110
Lys Ser Phe Gly Met Arg Gly Ser Glu Gly Cys Ala Tyr Leu Ser Asp
          115            120            125
Asp Gly Ala Ala Thr Ala Arg Cys Tyr Thr Glu Arg Lys Trp Ile Cys
          130            135            140
Arg Lys Arg Ile His
145

```

Sequence No.: 12

Sequence length: 188

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10122

Sequence description

```

Met Ser Thr Met Phe Ala Asp Thr Leu Leu Ile Val Phe Ile Ser Val
 1              5              10              15
Cys Thr Ala Leu Leu Ala Glu Gly Ile Thr Trp Val Leu Val Tyr Arg
      20              25              30
Thr Asp Lys Tyr Lys Arg Leu Lys Ala Glu Val Glu Lys Gln Ser Lys
      35              40              45
Lys Leu Glu Lys Lys Lys Glu Thr Ile Thr Glu Ser Ala Gly Arg Gln
      50              55              60
Gln Lys Lys Lys Ile Glu Arg Gln Glu Glu Lys Leu Lys Asn Asn Asn
      65              70              75              80
Arg Asp Leu Ser Met Val Arg Met Lys Ser Met Phe Ala Ile Gly Phe
      85              90              95
Cys Phe Thr Ala Leu Met Gly Met Phe Asn Ser Ile Phe Asp Gly Arg
      100              105              110
Val Val Ala Lys Leu Pro Phe Thr Pro Leu Ser Tyr Ile Gln Gly Leu
      115              120              125
Ser His Arg Asn Leu Leu Gly Asp Asp Thr Thr Asp Cys Ser Phe Ile
      130              135              140
Phe Leu Tyr Ile Leu Cys Thr Met Ser Ile Arg Gln Asn Ile Gln Lys
      145              150              155              160
Ile Leu Gly Leu Ala Pro Ser Arg Ala Ala Thr Lys Gln Ala Gly Gly
      165              170              175
Phe Leu Gly Pro Pro Pro Pro Ser Gly Lys Phe Ser
      180              185

```

Sequence No.: 13

Sequence length: 215

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Lymphoma

Cell line: U937

Clone name: HP10136

Sequence description

Met Val Leu Leu Thr Met Ile Ala Arg Val Ala Asp Gly Leu Pro Leu

105

1	5	10	15
Ala Ala Ser Met Gln Glu Asp Glu Gln Ser Gly Arg Asp Leu Gln Gln			
20	25	30	
Tyr Gln Ser Gln Ala Lys Gln Leu Phe Arg Lys Leu Asn Glu Gln Ser			
35	40	45	
Pro Thr Arg Cys Thr Leu Glu Ala Gly Ala Met Thr Phe His Tyr Ile			
50	55	60	
Ile Glu Gln Gly Val Cys Tyr Leu Val Leu Cys Glu Ala Ala Phe Pro			
65	70	75	80
Lys Lys Leu Ala Phe Ala Tyr Leu Glu Asp Leu His Ser Glu Phe Asp			
85	90	95	
Glu Gln His Gly Lys Lys Val Pro Thr Val Ser Arg Pro Tyr Ser Phe			
100	105	110	
Ile Glu Phe Asp Thr Phe Ile Gln Lys Thr Lys Lys Leu Tyr Ile Asp			
115	120	125	
Ser Arg Ala Arg Arg Asn Leu Gly Ser Ile Asn Thr Glu Leu Gln Asp			
130	135	140	
Val Gln Arg Ile Met Val Ala Asn Ile Glu Glu Val Leu Gln Arg Gly			
145	150	155	160
Glu Ala Leu Ser Ala Leu Asp Ser Lys Ala Asn Asn Leu Ser Ser Leu			
165	170	175	
Ser Lys Lys Tyr Arg Gln Asp Ala Lys Tyr Leu Asn Met Arg Ser Thr			
180	185	190	
Tyr Ala Lys Leu Ala Ala Val Ala Val Phe Phe Ile Met Leu Ile Val			
195	200	205	
Tyr Val Arg Phe Trp Trp Leu			
210	215		

Sequence No.: 14

Sequence length: 112

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10175

Sequence description

Met Gln Asp Thr Gly Ser Val Val Pro Leu His Trp Phe Gly Phe Gly
1 5 10 15
Tyr Ala Ala Leu Val Ala Ser Gly Gly Ile Ile Gly Tyr Val Lys Ala

106

20	25	30
Gly Ser Val Pro Ser Leu Ala Ala Gly Leu Leu Phe Gly Ser Leu Ala		
35	40	45
Gly Leu Gly Ala Tyr Gln Leu Ser Gln Asp Pro Arg Asn Val Trp Val		
50	55	60
Phe Leu Ala Thr Ser Gly Thr Leu Ala Gly Ile Met Gly Met Arg Phe		
65	70	75
Tyr His Ser Gly Lys Phe Met Pro Ala Gly Leu Ile Ala Gly Ala Ser		
85	90	95
Leu Leu Met Val Ala Lys Val Gly Val Ser Met Phe Asn Arg Pro His		
100	105	110

Sequence No.: 15

Sequence length: 114

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP10179

Sequence description

Met Glu Lys Pro Leu Phe Pro Leu Val Pro Leu His Trp Phe Gly Phe		
1	5	10
Gly Tyr Thr Ala Leu Val Val Ser Gly Gly Ile Val Gly Tyr Val Lys		
20	25	30
Thr Gly Ser Val Pro Ser Leu Ala Ala Gly Leu Leu Phe Gly Ser Leu		
35	40	45
Ala Gly Leu Gly Ala Tyr Gln Leu Tyr Gln Asp Pro Arg Asn Val Trp		
50	55	60
Gly Phe Leu Ala Ala Thr Ser Val Thr Phe Val Gly Val Met Gly Met		
65	70	75
Arg Ser Tyr Tyr Tyr Gly Lys Phe Met Pro Val Gly Leu Ile Ala Gly		
85	90	95
Ala Ser Leu Leu Met Ala Ala Lys Val Gly Val Arg Met Leu Met Thr		
100	105	110

Ser Asp

Sequence No.: 16



Sequence length: 327

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Fibrosarcoma

Cell line: HT-1080

Clone name: HP10196

Sequence description

```

Met Ala Ala Ala Ala Ala Ala Ala Ala Thr Asn Gly Thr Gly Gly
 1             5             10             15
Ser Ser Gly Met Glu Val Asp Ala Ala Val Val Pro Ser Val Met Ala
          20          25          30
Cys Gly Val Thr Gly Ser Val Ser Val Ala Leu His Pro Leu Val Ile
      35          40          45
Leu Asn Ile Ser Asp His Trp Ile Arg Met Arg Ser Gln Glu Gly Arg
      50          55          60
Pro Val Gln Val Ile Gly Ala Leu Ile Gly Lys Gln Glu Gly Arg Asn
      65          70          75          80
Ile Glu Val Met Asn Ser Phe Glu Leu Leu Ser His Thr Val Glu Glu
          85          90          95
Lys Ile Ile Ile Asp Lys Glu Tyr Tyr Tyr Thr Lys Glu Glu Gln Phe
      100          105          110
Lys Gln Val Phe Lys Glu Leu Glu Phe Leu Gly Trp Tyr Thr Thr Gly
      115          120          125
Gly Pro Pro Asp Pro Ser Asp Ile His Val His Lys Gln Val Cys Glu
      130          135          140
Ile Ile Glu Ser Pro Leu Phe Leu Lys Leu Asn Pro Met Thr Lys His
      145          150          155          160
Thr Asp Leu Pro Val Ser Val Phe Glu Ser Val Ile Asp Ile Ile Asn
          165          170          175
Gly Glu Ala Thr Met Leu Phe Ala Glu Leu Thr Tyr Thr Leu Ala Thr
          180          185          190
Glu Glu Ala Glu Arg Ile Gly Val Asp His Val Ala Arg Met Thr Ala
          195          200          205
Thr Gly Ser Gly Glu Asn Ser Thr Val Ala Glu His Leu Ile Ala Gln
          210          215          220
His Ser Ala Ile Lys Met Leu His Ser Arg Val Lys Leu Ile Leu Glu
      225          230          235          240
Tyr Val Lys Ala Ser Glu Ala Gly Glu Val Pro Phe Asn His Glu Ile
          245          250          255

```

```

Leu Arg Glu Ala Tyr Ala Leu Cys His Cys Leu Pro Val Leu Ser Thr
      260                      265                      270
Asp Lys Phe Lys Thr Asp Phe Tyr Asp Gln Cys Asn Asp Val Gly Leu
      275                      280                      285
Met Ala Tyr Leu Gly Thr Ile Thr Lys Thr Cys Asn Thr Met Asn Gln
      290                      295                      300
Phe Val Asn Lys Phe Asn Val Leu Tyr Asp Arg Gln Gly Ile Gly Arg
      305                      310                      315                      320
Arg Met Arg Gly Leu Phe Phe
      325

```

Sequence No.: 17

Sequence length: 373

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Fibrosarcoma

Cell line: HT-1080

Clone name: HP10235

Sequence description

```

Met Thr Leu Cys Ala Met Leu Pro Leu Leu Leu Phe Thr Tyr Leu Asn
 1              5              10              15
Ser Phe Leu His Gln Arg Ile Pro Gln Ser Val Arg Ile Leu Gly Ser
      20              25              30
Leu Val Ala Ile Leu Leu Val Phe Leu Ile Thr Ala Ile Leu Val Lys
      35              40              45
Val Gln Leu Asp Ala Leu Pro Phe Phe Val Ile Thr Met Ile Lys Ile
      50              55              60
Val Leu Ile Asn Ser Phe Gly Ala Ile Leu Gln Gly Ser Leu Phe Gly
      65              70              75              80
Leu Ala Gly Leu Leu Pro Ala Ser Tyr Thr Ala Pro Ile Met Ser Gly
      85              90              95
Gln Gly Leu Ala Gly Phe Phe Ala Ser Val Ala Met Ile Cys Ala Ile
      100             105             110
Ala Ser Gly Ser Glu Leu Ser Glu Ser Ala Phe Gly Tyr Phe Ile Thr
      115             120             125
Ala Cys Ala Val Ile Ile Leu Thr Ile Ile Cys Tyr Leu Gly Leu Pro
      130             135             140
Arg Leu Glu Phe Tyr Arg Tyr Tyr Gln Gln Leu Lys Leu Glu Gly Pro

```

145	150	155	160
Gly Glu Gln Glu Thr Lys Leu Asp Leu Ile Ser Lys Gly Glu Glu Pro			
	165	170	175
Arg Ala Gly Lys Glu Glu Ser Gly Val Ser Val Ser Asn Ser Gln Pro			
	180	185	190
Thr Asn Glu Ser His Ser Ile Lys Ala Ile Leu Lys Asn Ile Ser Val			
	195	200	205
Leu Ala Phe Ser Val Cys Phe Ile Phe Thr Ile Thr Ile Gly Met Phe			
	210	215	220
Pro Ala Val Thr Val Glu Val Lys Ser Ser Ile Ala Gly Ser Ser Thr			
	225	230	235
Trp Glu Arg Tyr Phe Ile Pro Val Ser Cys Phe Leu Thr Phe Asn Ile			
	245	250	255
Phe Asp Trp Leu Gly Arg Ser Leu Thr Ala Val Phe Met Trp Pro Gly			
	260	265	270
Lys Asp Ser Arg Trp Leu Pro Ser Leu Val Leu Ala Arg Leu Val Phe			
	275	280	285
Val Pro Leu Leu Leu Leu Cys Asn Ile Lys Pro Arg Arg Tyr Leu Thr			
	290	295	300
Val Val Phe Glu His Asp Ala Trp Phe Ile Phe Phe Met Ala Ala Phe			
	305	310	315
Ala Phe Ser Asn Gly Tyr Leu Ala Ser Leu Cys Met Cys Phe Gly Pro			
	325	330	335
Lys Lys Val Lys Pro Ala Glu Ala Glu Thr Ala Gly Ala Ile Met Ala			
	340	345	350
Phe Phe Leu Cys Leu Gly Leu Ala Leu Gly Ala Val Phe Ser Phe Leu			
	355	360	365
Phe Arg Ala Ile Val			
	370		

Sequence No.: 18

Sequence length: 183

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10297

Sequence description

110

```

Met Lys Leu Leu Ser Leu Val Ala Val Val Gly Cys Leu Leu Val Pro
 1              5              10              15
Pro Ala Glu Ala Asn Lys Ser Ser Glu Asp Ile Arg Cys Lys Cys Ile
      20              25              30
Cys Pro Pro Tyr Arg Asn Ile Ser Gly His Ile Tyr Asn Gln Asn Val
      35              40              45
Ser Gln Lys Asp Cys Asn Cys Leu His Val Val Glu Pro Met Pro Val
      50              55              60
Pro Gly His Asp Val Glu Ala Tyr Cys Leu Leu Cys Glu Cys Arg Tyr
      65              70              75              80
Glu Glu Arg Ser Thr Thr Thr Ile Lys Val Ile Ile Val Ile Tyr Leu
      85              90              95
Ser Val Val Gly Ala Leu Leu Leu Tyr Met Ala Phe Leu Met Leu Val
      100              105              110
Asp Pro Leu Ile Arg Lys Pro Asp Ala Tyr Thr Glu Gln Leu His Asn
      115              120              125
Glu Glu Glu Asn Glu Asp Ala Arg Ser Met Ala Ala Ala Ala Ala Ser
      130              135              140
Leu Gly Gly Pro Arg Ala Asn Thr Val Leu Glu Arg Val Glu Gly Ala
      145              150              155              160
Gln Gln Arg Trp Lys Leu Gln Val Gln Glu Gln Arg Lys Thr Val Phe
      165              170              175
Asp Arg His Lys Met Leu Ser
      180

```

Sequence No.: 19

Sequence length: 116

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10299

Sequence description

```

Met Ala Ser Thr Val Val Ala Val Gly Leu Thr Ile Ala Ala Ala Gly
 1              5              10              15
Phe Ala Gly Arg Tyr Val Leu Gln Ala Met Lys His Met Glu Pro Gln
      20              25              30
Val Lys Gln Val Phe Gln Ser Leu Pro Lys Ser Ala Phe Ser Gly Gly
      35              40              45

```

111

Tyr Tyr Arg Gly Gly Phe Glu Pro Lys Met Thr Lys Arg Glu Ala Ala  
 50 55 60  
 Leu Ile Leu Gly Val Ser Pro Thr Ala Asn Lys Gly Lys Ile Arg Asp  
 65 70 75 80  
 Ala His Arg Arg Ile Met Leu Leu Asn His Pro Asp Lys Gly Gly Ser  
 85 90 95  
 Pro Tyr Ile Ala Ala Lys Ile Asn Glu Ala Lys Asp Leu Leu Glu Gly  
 100 105 110  
 Gln Ala Lys Lys  
 115

Sequence No.: 20

Sequence length: 152

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP10301

Sequence description

Met Ala Val Leu Ser Lys Glu Tyr Gly Phe Val Leu Leu Thr Gly Ala  
 1 5 10 15  
 Ala Ser Phe Ile Met Val Ala His Leu Ala Ile Asn Val Ser Lys Ala  
 20 25 30  
 Arg Lys Lys Tyr Lys Val Glu Tyr Pro Ile Met Tyr Ser Thr Asp Pro  
 35 40 45  
 Glu Asn Gly His Ile Phe Asn Cys Ile Gln Arg Ala His Gln Asn Thr  
 50 55 60  
 Leu Glu Val Tyr Pro Pro Phe Leu Phe Phe Leu Ala Val Gly Gly Val  
 65 70 75 80  
 Tyr His Pro Arg Ile Ala Ser Gly Leu Gly Leu Ala Trp Ile Val Gly  
 85 90 95  
 Arg Val Leu Tyr Ala Tyr Gly Tyr Tyr Thr Gly Glu Pro Ser Lys Arg  
 100 105 110  
 Ser Arg Gly Ala Leu Gly Ser Ile Ala Leu Leu Gly Leu Val Gly Thr  
 115 120 125  
 Thr Val Cys Ser Ala Phe Gln His Leu Gly Trp Val Lys Ser Gly Leu  
 130 135 140  
 Gly Ser Gly Pro Lys Cys Cys His

145

150

Sequence No.: 21

Sequence length: 559

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Liver

Clone name: HP10302

Sequence description

```

Met Ala Pro Thr Leu Gln Gln Ala Tyr Arg Arg Arg Trp Trp Met Ala
 1             5             10             15
Cys Thr Ala Val Leu Glu Asn Leu Phe Phe Ser Ala Val Leu Leu Gly
      20             25             30
Trp Gly Ser Leu Leu Ile Ile Leu Lys Asn Glu Gly Phe Tyr Ser Ser
      35             40             45
Thr Cys Pro Ala Glu Ser Ser Thr Asn Thr Thr Gln Asp Glu Gln Arg
      50             55             60
Arg Trp Pro Gly Cys Asp Gln Gln Asp Glu Met Leu Asn Leu Gly Phe
      65             70             75             80
Thr Ile Gly Ser Phe Val Leu Ser Ala Thr Thr Leu Pro Leu Gly Ile
      85             90             95
Leu Met Asp Arg Phe Gly Pro Arg Pro Val Arg Leu Val Gly Ser Ala
      100            105            110
Cys Phe Thr Ala Ser Cys Thr Leu Met Ala Leu Ala Ser Arg Asp Val
      115            120            125
Glu Ala Leu Ser Pro Leu Ile Phe Leu Ala Leu Ser Leu Asn Gly Phe
      130            135            140
Gly Gly Ile Cys Leu Thr Phe Thr Ser Leu Thr Leu Pro Asn Met Phe
      145            150            155            160
Gly Asn Leu Arg Ser Thr Leu Met Ala Leu Met Ile Gly Ser Tyr Ala
      165            170            175
Ser Ser Ala Ile Thr Phe Pro Gly Ile Lys Leu Ile Tyr Asp Ala Gly
      180            185            190
Val Ala Phe Val Val Ile Met Phe Thr Trp Ser Gly Leu Ala Cys Leu
      195            200            205
Ile Phe Leu Asn Cys Thr Leu Asn Trp Pro Ile Glu Ala Phe Pro Ala
      210            215            220
Pro Glu Glu Val Asn Tyr Thr Lys Lys Ile Lys Leu Ser Gly Leu Ala

```

225	230	235	240
Leu Asp His Lys Val Thr Gly Asp Leu Phe Tyr Thr His Val Thr Thr			
	245	250	255
Met Gly Gln Arg Leu Ser Gln Lys Ala Pro Ser Leu Glu Asp Gly Ser			
	260	265	270
Asp Ala Phe Met Ser Pro Gln Asp Val Arg Gly Thr Ser Glu Asn Leu			
	275	280	285
Pro Glu Arg Ser Val Pro Leu Arg Lys Ser Leu Cys Ser Pro Thr Phe			
	290	295	300
Leu Trp Ser Leu Leu Thr Met Gly Met Thr Gln Leu Arg Ile Ile Phe			
305	310	315	320
Tyr Met Ala Ala Val Asn Lys Met Leu Glu Tyr Leu Val Thr Gly Gly			
	325	330	335
Gln Glu His Glu Thr Asn Glu Gln Gln Gln Lys Val Ala Glu Thr Val			
	340	345	350
Gly Phe Tyr Ser Ser Val Phe Gly Ala Met Gln Leu Leu Cys Leu Leu			
	355	360	365
Thr Cys Pro Leu Ile Gly Tyr Ile Met Asp Trp Arg Ile Lys Asp Cys			
	370	375	380
Val Asp Ala Pro Thr Gln Gly Thr Val Leu Gly Asp Ala Arg Asp Gly			
385	390	395	400
Val Ala Thr Lys Ser Ile Arg Pro Arg Tyr Cys Lys Ile Gln Lys Leu			
	405	410	415
Thr Asn Ala Ile Ser Ala Phe Thr Leu Thr Asn Leu Leu Leu Val Gly			
	420	425	430
Phe Gly Ile Thr Cys Leu Ile Asn Asn Leu His Leu Gln Phe Val Thr			
	435	440	445
Phe Val Leu His Thr Ile Val Arg Gly Phe Phe His Ser Ala Cys Gly			
	450	455	460
Ser Leu Tyr Ala Ala Val Phe Pro Ser Asn His Phe Gly Thr Leu Thr			
465	470	475	480
Gly Leu Gln Ser Leu Ile Ser Ala Val Phe Ala Leu Leu Gln Gln Pro			
	485	490	495
Leu Phe Met Ala Met Val Gly Pro Leu Lys Gly Glu Pro Phe Trp Val			
	500	505	510
Asn Leu Gly Leu Leu Leu Phe Ser Leu Leu Gly Phe Leu Leu Pro Ser			
	515	520	525
Tyr Leu Phe Tyr Tyr Arg Ala Arg Leu Gln Gln Glu Tyr Ala Ala Asn			
	530	535	540
Gly Met Gly Pro Leu Lys Val Leu Ser Gly Ser Glu Val Thr Ala			
545	550	555	

Sequence length: 330

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Osteosarcoma

Cell line: U-2 OS

Clone name: HP10304

Sequence description

Met	Glu	Gly	Ala	Pro	Pro	Gly	Ser	Leu	Ala	Leu	Arg	Leu	Leu	Leu	Phe	1	5	10	15
Val	Ala	Leu	Pro	Ala	Ser	Gly	Trp	Leu	Thr	Thr	Gly	Ala	Pro	Glu	Pro	20	25	30	
Pro	Pro	Leu	Ser	Gly	Ala	Pro	Gln	Asp	Gly	Ile	Arg	Ile	Asn	Val	Thr	35	40	45	
Thr	Leu	Lys	Asp	Asp	Gly	Asp	Ile	Ser	Lys	Gln	Gln	Val	Val	Leu	Asn	50	55	60	
Ile	Thr	Tyr	Glu	Ser	Gly	Gln	Val	Tyr	Val	Asn	Asp	Leu	Pro	Val	Asn	65	70	75	80
Ser	Gly	Val	Thr	Arg	Ile	Ser	Cys	Gln	Thr	Leu	Ile	Val	Lys	Asn	Glu	85	90	95	
Asn	Leu	Glu	Asn	Leu	Glu	Glu	Lys	Glu	Tyr	Phe	Gly	Ile	Val	Ser	Val	100	105	110	
Arg	Ile	Leu	Val	His	Glu	Trp	Pro	Met	Thr	Ser	Gly	Ser	Ser	Leu	Gln	115	120	125	
Leu	Ile	Val	Ile	Gln	Glu	Glu	Val	Val	Glu	Ile	Asp	Gly	Lys	Gln	Val	130	135	140	
Gln	Gln	Lys	Asp	Val	Thr	Glu	Ile	Asp	Ile	Leu	Val	Lys	Asn	Arg	Gly	145	150	155	160
Val	Leu	Arg	His	Ser	Asn	Tyr	Thr	Leu	Pro	Leu	Glu	Glu	Ser	Met	Leu	165	170	175	
Tyr	Ser	Ile	Ser	Arg	Asp	Ser	Asp	Ile	Leu	Phe	Thr	Leu	Pro	Asn	Leu	180	185	190	
Ser	Lys	Lys	Glu	Ser	Val	Ser	Ser	Leu	Gln	Thr	Thr	Ser	Gln	Tyr	Leu	195	200	205	
Ile	Arg	Asn	Val	Glu	Thr	Thr	Val	Asp	Glu	Asp	Val	Leu	Pro	Gly	Lys	210	215	220	
Leu	Pro	Glu	Thr	Pro	Leu	Arg	Ala	Glu	Pro	Pro	Ser	Ser	Tyr	Lys	Val	225	230	235	240
Met	Cys	Gln	Trp	Met	Glu	Lys	Phe	Arg	Lys	Asp	Leu	Cys	Arg	Phe	Trp	245	250	255	



115

Ser Asn Val Phe Pro Val Phe Phe Gln Phe Leu Asn Ile Met Val Val  
 260 265 270  
 Gly Ile Thr Gly Ala Ala Val Val Ile Thr Ile Leu Lys Val Phe Phe  
 275 280 285  
 Pro Val Ser Glu Tyr Lys Gly Ile Leu Gln Leu Asp Lys Val Asp Val  
 290 295 300  
 Ile Pro Val Thr Ala Ile Asn Leu Tyr Pro Asp Gly Pro Glu Lys Arg  
 305 310 315 320  
 Ala Glu Asn Leu Glu Asp Lys Thr Cys Ile  
 325 330

Sequence No.: 23

Sequence length: 108

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Osteosarcoma

Cell line: HU-2 OS

Clone name: HP10305

Sequence description

Met Ser Leu Thr Ser Ser Ser Ser Val Arg Val Glu Trp Ile Ala Ala  
 1 5 10 15  
 Val Thr Ile Ala Ala Gly Thr Ala Ala Ile Gly Tyr Leu Ala Tyr Lys  
 20 25 30  
 Arg Phe Tyr Val Lys Asp His Arg Asn Lys Ala Met Ile Asn Leu His  
 35 40 45  
 Ile Gln Lys Asp Asn Pro Lys Ile Val His Ala Phe Asp Met Glu Asp  
 50 55 60  
 Leu Gly Asp Lys Ala Val Tyr Cys Arg Cys Trp Arg Ser Lys Lys Phe  
 65 70 75 80  
 Pro Phe Cys Asp Gly Ala His Thr Lys His Asn Glu Glu Thr Gly Asp  
 85 90 95  
 Asn Val Gly Pro Leu Ile Ile Lys Lys Lys Glu Thr  
 100 105

Sequence No.: 24

Sequence length: 101

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Osteosarcoma

Cell line: U-2 OS

Clone name: HP10306

Sequence description

Met Asn Leu Glu Arg Val Ser Asn Glu Glu Lys Leu Asn Leu Cys Arg  
1 5 10 15  
Lys Tyr Tyr Leu Gly Gly Phe Ala Phe Leu Pro Phe Leu Trp Leu Val  
20 25 30  
Asn Ile Phe Trp Phe Phe Arg Glu Ala Phe Leu Val Pro Ala Tyr Thr  
35 40 45  
Glu Gln Ser Gln Ile Lys Gly Tyr Val Trp Arg Ser Ala Val Gly Phe  
50 55 60  
Leu Phe Trp Val Ile Val Leu Thr Ser Trp Ile Thr Ile Phe Gln Ile  
65 70 75 80  
Tyr Arg Pro Arg Trp Gly Ala Leu Gly Asp Tyr Leu Ser Phe Thr Ile  
85 90 95  
Pro Leu Gly Thr Pro  
100

Sequence No.: 25

Sequence length: 372

Sequence type: Amino acid

Topology: Linear

Sequence kind: Protein

Hypothetical: No

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP10328

Sequence description

Met Lys Tyr Leu Arg His Arg Arg Pro Asn Ala Thr Leu Ile Leu Ala  
1 5 10 15  
Ile Gly Ala Phe Thr Leu Leu Leu Phe Ser Leu Leu Val Ser Pro Pro  
20 25 30  
Thr Cys Lys Val Gln Glu Gln Pro Pro Ala Ile Pro Glu Ala Leu Ala

35	40	45
Trp Pro Thr Pro Pro Thr Arg Pro Ala Pro Ala Pro Cys His Ala Asn		
50	55	60
Thr Ser Met Val Thr His Pro Asp Phe Ala Thr Gln Pro Gln His Val		
65	70	75
Gln Asn Phe Leu Leu Tyr Arg His Cys Arg His Phe Pro Leu Leu Gln		
85	90	95
Asp Val Pro Pro Ser Lys Cys Ala Gln Pro Val Phe Leu Leu Leu Val		
100	105	110
Ile Lys Ser Ser Pro Ser Asn Tyr Val Arg Arg Glu Leu Leu Arg Arg		
115	120	125
Thr Trp Gly Arg Glu Arg Lys Val Arg Gly Leu Gln Leu Arg Leu Leu		
130	135	140
Phe Leu Val Gly Thr Ala Ser Asn Pro His Glu Ala Arg Lys Val Asn		
145	150	155
Arg Leu Leu Glu Leu Glu Ala Gln Thr His Gly Asp Ile Leu Gln Trp		
165	170	175
Asp Phe His Asp Ser Phe Phe Asn Leu Thr Leu Lys Gln Val Leu Phe		
180	185	190
Leu Gln Trp Gln Glu Thr Arg Cys Ala Asn Ala Ser Phe Val Leu Asn		
195	200	205
Gly Asp Asp Asp Val Phe Ala His Thr Asp Asn Met Val Phe Tyr Leu		
210	215	220
Gln Asp His Asp Pro Gly Arg His Leu Phe Val Gly Gln Leu Ile Gln		
225	230	235
Asn Val Gly Pro Ile Arg Ala Phe Trp Ser Lys Tyr Tyr Val Pro Glu		
245	250	255
Val Val Thr Gln Asn Glu Arg Tyr Pro Pro Tyr Cys Gly Gly Gly Gly		
260	265	270
Phe Leu Leu Ser Arg Phe Thr Ala Ala Ala Leu Arg Arg Ala Ala His		
275	280	285
Val Leu Asp Ile Phe Pro Ile Asp Asp Val Phe Leu Gly Met Cys Leu		
290	295	300
Glu Leu Glu Gly Leu Lys Pro Ala Ser His Ser Gly Ile Arg Thr Ser		
305	310	315
Gly Val Arg Ala Pro Ser Gln His Leu Ser Ser Phe Asp Pro Cys Phe		
325	330	335
Tyr Arg Asp Leu Leu Leu Val His Arg Phe Leu Pro Tyr Glu Met Leu		
340	345	350
Leu Met Trp Asp Ala Leu Asn Gln Pro Asn Leu Thr Cys Gly Asn Gln		
355	360	365
Thr Gln Ile Tyr		
370		

Sequence No.: 26  
Sequence length: 615  
Sequence type: Nucleic acid  
Strandedness: Double  
Topology: Linear  
Sequence kind: cDNA to mRNA  
Original source:  
Organism species: *Homo sapiens*  
Cell kind: Fibrosarcoma  
Cell line: HT-1080  
Clone name: HP00442  
Sequence description

ATGACGGGGC TAGCACTGCT CTACTCCGGG GTCTTCGTGG CCTTCTGGGC CTGCGCGCTG	60
GGCGTGGGAG TCTGCTACAC CATTTTGTAT TTGGGCTTCC GCTTTGATGT GGCATGGTTC	120
CTGACGGAGA CTTGCCCCTT CATGTGGTCC AACCTGGGCA TTGGCCTAGC TATCTCCCTG	180
TCTGTGGTTG GGGCAGCCTG GGGCATCTAT ATTACCGGCT CCTCCATCAT TGGTGGAGGA	240
GTGAAGGCCC CCAGGATCAA GACCAAGAAC CTGGTCAGCA TCATCTTCTG TGAGGCTGTG	300
GCCATCTACG GCATCATCAT GGCAATTGTC ATTAGCAACA TGGCTGAGCC TTTCACTGCC	360
ACAGACCCCA AGGCCATCGG CCATCGGAAC TACCATGCAG GCTACTCCAT GTTTGGGGCT	420
GGCCTCACCG TAGGCCTGTC TAACCTCTTC TGTGGAGTCT GCGTGGGCAT CGTGGGCAGT	480
GGGGCTGCCC TGGCCGATGC TCAGAACCCC AGCCTCTTTG TAAAGATTCT CATCGTGGAG	540
ATCTTTGGCA GCGCCATTGG CCTCTTTGGG GTCATCGTGC CAATTCTTCA GACCTCCAGA	600
GTGAAGATGG GTGAC	615

Sequence No.: 27  
Sequence length: 1113  
Sequence type: Nucleic acid  
Strandedness: Double  
Topology: Linear  
Sequence kind: cDNA to mRNA  
Original source:  
Organism species: *Homo sapiens*  
Cell kind: Leukocyte  
Clone name: HP00804  
Sequence description

ATGTCCCATG AAAAGAGTTT TTTGGTGTCT GGGGACAACT ATCCTCCCCC CAACCCTGGA	60
TATCCGGGGG GGCCCCAGCC ACCCATGCCC CCCTATGCTC AGCCTCCCTA CCCTGGGGCC	120
CCTTACCCAC AGCCCCCTTT CCAGCCCTCC CCCTACGGTC AGCCAGGGTA CCCCCATGGC	180
CCCAGCCCCCT ACCCCCAAGG GGGCTACCCA CAGGGTCCCT ACCCCCAAGG GGGCTACCCA	240
CAGGGCCCCCT ACCCACAAGA GGGCTACCCA CAGGGCCCCCT ACCCCCAAGG GGGCTACCCC	300

CAGGGGCCAT ATCCCCAGAG CCCCTTCCCC CCCAACCCCT ATGGACAGCC ACAGGTCTTC	360
CCAGGACAAG ACCCTGACTC ACCCCAGCAT GGAAACTACC AGGAGGAGGG TCCCCATCC	420
TACTATGACA ACCAGGACTT CCCTGCCACC AACTGGGATG ACAAGAGCAT CCGACAGGCC	480
TTCATCCGCA AGGTGTTCTT AGTGCTGACC TTGCAGCTGT CGGTGACCCT GTCCACGGTG	540
TCTGTGTTCA CTTTGTGTC GGAGGTGAAG GGCTTGTCC GGGAGAATGT CTGGACCTAC	600
TATGTCTCCT ATGCTGTCTT CTTTCTCTCT CTCATCGTCC TCAGCTGTTG TGGGGACTTC	660
CGGCGAAAGC ACCCCTGGAA CCTTGTGCA CTGTGGGTCC TGACCGCCAG CCTGTCGTAC	720
ATGGTGGGA TGATCGCCAG CTTCTACAAC ACCGAGGCAG TCATCATGGC CGTGGGCATC	780
ACCACAGCCG TCTGCTTCAC CGTCGTATC TTCTCCATGC AGACCCGCTA CGACTTCACC	840
TCATGCATGG GCGTGCTCCT GGTGAGCATG GTGGTGCTCT TCATCTTCGC CATTCTCTGC	900
ATCTTCATCC GGAACCGCAT CCTGGAGATC GTGTACGCTT CACTGGGCGC TCTGCTCTTC	960
ACCTGCTTCC TCGCAGTGGA CACCCAGCTG CTGTGGGGA ACAAGCAGCT GTCCCTGAGC	1020
CCAGAAGAGT ATGTGTTTGC TGGGCTGAAC CTGTACACAG ACATCATCAA CATCTTCCTG	1080
TACATCTCA CCATCATTGG CCGCGCCAAG GAG	1113

Sequence No.: 28

Sequence length: 537

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP01098

Sequence description

ATGCTGTCTC TAGACTTTTT GGACGATGTG CGGCGGATGA ACAAGCGGCA GCTCTATTAT	60
CAAGTCCTAA ATTTTGAAT GATTGTCTCA TCGGCACTAA TGATCTGGAA GGGGTTAATG	120
GTAATAACTG GAAGTGAAG TCCGATTGTA GTGGTGCTCA GTGGCAGCAT GGAACCTGCA	180
TTTCATAGAG GAGATCTTCT CTTTCTAACA AATCGAGTTG AAGATCCCAT ACGAGTGCGA	240
GAAATTGTTG TTTTtaggat AGAAGGAAGA GAGATTCCTA TAGTTCACCG AGTCTTGAAG	300
ATTCATGAAA AGCAAAATGG GCATATCAAG TTTTtgacca AAGGAGATAA TAATGCGGTT	360
GATGACCGAG GCCTCTATAA ACAAGGACAA CATTGGCTAG AGAAAAAAGA TGTGTGGGG	420
AGAGCCAGGG GATTTGTTCC TTATATTGGA ATTGTGACGA TCCTCATGAA TGAATATCCT	480
AAATTTAAGT ATGCAGTTCT CTTTTGCTG GGTttattcg TGCTGGTTCA TCGTGAG	537

Sequence No.: 29

Sequence length: 1041

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Liver

Clone name: HP01148

Sequence description

```
ATGGCTCTGC TATTCTCCTT GATCCTTGCC ATTTGCACCA GACCTGGATT CCTAGCGTCT    60
CCATCTGGAG TGCGGCTGGT GGGGGGCCTC CACCGCTGTG AAGGGCGGGT GGAGGTGGAA    120
CAGAAAGGCC AGTGGGGCAC CGTGTGTGAT GACGGCTGGG ACATTAAGGA CGTGGCTGTG    180
TTGTGCCGGG AGCTGGGCTG TGGAGCTGCC AGCGGAACCC CTAGTGGTAT TTTGTATGAG    240
CCACCAGCAG AAAAAGAGCA AAAGGTCCTC ATCCAATCAG TCAGTTGCAC AGGAACAGAA    300
GATACATTGG CTCAGTGTGA GCAAGAAGAA GTTTATGATT GTTCACATGA AGAAGATGCT    360
GGGGCATCGT GTGAGAACCC AGAGAGCTCT TTCTCCCCAG TCCCAGAGGG TGTCAAGGCTG    420
GCTGACGGCC CTGGGCATTG CAAGGGACGC GTGGAAGTGA AGCACCAGAA CCAGTGGTAT    480
ACCGTGTGCC AGACAGGCTG GAGCCTCCGG GCCGCAAAGG TGGTGTGCCG GCAGCTGGGA    540
TGTGGGAGGG CTGTACTGAC TCAAAAACGC TGCAACAAGC ATGCCTATGG CCGAAAACCC    600
ATCTGGCTGA GCCAGATGTC ATGCTCAGGA CGAGAAGCAA CCCTTCAGGA TTGCCCTTCT    660
GGGCCTTGGG GGAAGAACAC CTGCAACCAT GATGAAGACA CGTGGGTCGA ATGTGAAGAT    720
CCCTTTGACT TGAGACTAGT AGGAGGAGAC AACCTCTGCT CTGGGCGACT GGAGGTGCTG    780
CACAAGGGCG TATGGGGCTC TGTCTGTGAT GACAACTGGG GAGAAAAGGA GGACCAGGTC    840
GTATGCAAGC AACTGGGCTG TGGGAAGTCC CTCTCTCCCT CCTTCAGAGA CCGGAAATGC    900
TATGGCCCTG GGGTTGGCCG CATCTGGCTG GATAATGTTT GTTGCTCAGG GGAGGAGCAG    960
TCCCTGGAGC AGTGCCAGCA CAGATTTTGG GGGTTTCACG ACTGCACCCA CCAGGAAGAT   1020
GTGGCTGTCA TCTGCTCAGG A                               1041
```

Sequence No.: 30

Sequence length: 1662

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Liver

Clone name: HP01293

Sequence description

```
ATGCCCACCG TGGATGACAT TCTGGAGCAG GTTGGGGAGT CTGGCTGGTT CCAGAAGCAA    60
GCCTTCCTCA TCTTATGCCT GCTGTGGGCT GCCTTTGCGC CCATCTGTGT GGGCATCGTC    120
TTCCTGGGTT TCACACCTGA CCACCACTGC CAGAGTCCTG GGGTGGCTGA GCTGAGCCAG    180
CGCTGTGGCT GGAGCCCTGC GGAGGAGCTG AACTATACAG TGCCAGGCCT GGGGCCCGCG    240
GGCGAGGCCT TCCTTGGCCA GTGCAGGCGC TATGAAGTGG ACTGGAACCA GAGCGCCCTC    300
```

AGCTGTGTAG	ACCCCTGGC	TAGCCTGGCC	ACCAACAGGA	GCCACCTGCC	GCTGGGTCCC	360
TGCCAGGATG	GCTGGGTGTA	TGACACGCCC	GGCTCTTCCA	TGCTCACTGA	GTTCAACCTG	420
GTGTGTGCTG	ACTCCTGGAA	GCTGGACCTC	TTTCAGTCCT	GTTTGAATGC	GGGCTTCTTC	480
TTTGGCTCTC	TGGGTGTTGG	CTACTTTGCA	GACAGGTTTG	GCCGTAAGCT	GTGTCTCCTG	540
GGAAGTGTGC	TGGTCAACGC	GGTGTGCGGC	GTGTCATGG	CCTTCTCGCC	CAACTACATG	600
TCCATGCTGC	TCTTCCGCCT	GCTGCAGGGC	CTGGTCAGCA	AGGGCAACTG	GATGGCTGGC	660
TACACCTTAA	TCACAGAATT	TGTTGGCTCG	GGCTCCAGAA	GAACGGTGGC	GATCATGTAC	720
CAGATGGCCT	TCACGGTGGG	GCTGGTGGCG	CTTACCGGGC	TGGCCTACGC	CCTGCCTCAC	780
TGGCGCTGGC	TGCAGCTGGC	AGTCTCCCTG	CCCACCTTCC	TCTTCCTGCT	CTACTACTGG	840
TGTGTGCCGG	AGTCCCCTCG	GTGGCTGTTA	TCACAAAAAA	GAAACACTGA	AGCAATAAAG	900
ATAATGGACC	ACATCGCTCA	AAAGAATGGG	AAGTTGCCCTC	CTGCTGATTT	AAAGATGCTT	960
TCCCTCGAAG	AGGATGTCAC	CGAAAAGCTG	AGCCCTTCAT	TTGCAGACCT	GTTCCGCACG	1020
CCGCGCCTGA	GGAAGCGCAC	CTTCATCCTG	ATGTACCTGT	GGTTCACGGA	CTCTGTGCTC	1080
TATCAGGGGC	TCATCCTGCA	CATGGGCGCC	ACCAGCGGGA	ACCTCTACCT	GGATTTCCCTT	1140
TACTCCGCTC	TGGTCGAAAT	CCCGGGGGCC	TTCATAGCCC	TCATCACCAT	TGACCGCGTG	1200
GGCCGCATCT	ACCCCATGGC	CGTGTCAAAT	TTGTTGGCGG	GGGCAGCCTG	CCTCGTCATG	1260
ATTTTTATCT	CACCTGACCT	GCACTGGTTA	AACATCATAA	TCATGTGTGT	TGGCCGAATG	1320
GGAATCACCA	TTGCAATACA	AATGATCTGC	CTGGTGAATG	CTGAGCTGTA	CCCCACATTC	1380
GTCAGGAACC	TGGAGTGAT	GGTGTGTTCC	TCCCTGTGTG	ACATAGGTGG	GATAATCACC	1440
CCCTTCATAG	TCTTCAGCCT	GAGGGAGGTC	TGGCAAGCCT	TGCCCCTCAT	TTTGTTCGCG	1500
GTGTTGGGCC	TGCTTGCCGC	GGGAGTGACG	CTACTTCTTC	CAGAGACCAA	GGGGGTGCGT	1560
TTGCCAGAGA	CCATGAAGGA	CGCCGAGAAC	CTTGGGAGAA	AAGCAAAGCC	CAAAGAAAAC	1620
ACGATTTACC	TTAAGGTCCA	AACCTCAGAA	CCCTCGGGCA	CC		1662

Sequence No.: 31

Sequence length: 1050

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP10013

Sequence description

ATGGCTGTGT	TTGTCGTGCT	CCTGGCGTTG	GTGGCGGGTG	TTTGGGGGAA	CGAGTTTAGT	60
ATATTAAAAT	CACCAGGGTC	TGTTGTTTTT	CGAAATGGAA	ATTGGCCTAT	ACCAGGAGAG	120
CGGATCCCAG	ACGTGGCTGC	ATTGTCCATG	GGCTTCTCTG	TGAAAGAAGA	CCTTTCTTGG	180
CCAGGACTCG	CAGTGGGTAA	CCTGTTTCAT	CGTCCTCGGG	CTACCGTCAT	GGTGATGGTG	240
AAGGGAGTGA	ACAAACTGGC	TCTACCCCCA	GGCAGTGTC	TTTCGTACCC	TTTGGAGAAT	300
GCAGTTCCTT	TTAGTCTTGA	CAGTGTGCA	AATTCCATTC	ACTCCTTATT	TTCTGAGGAA	360

ACTCCTGTTG	TTTTGCAGTT	GGCTCCCAGT	GAGGAAAGAG	TGTATATGCT	AGGGAAGGCA	420
AACTCAGTGT	TTGAAGACCT	TTCAGTCACC	TTGCGCCAGC	TCCGTAATCG	CCTGTTTCAA	480
GAAAACTCTG	TTCTCAGTTC	ACTCCCCCTC	AATTCTCTGA	GTAGGAACAA	TGAAGTTGAC	540
CTGCTCTTTC	TTTCTGAACT	GCAAGTGCTA	CATGATATTT	CAAGCTTGCT	GTCTCGTCAT	600
AAGCATCTAG	CCAAGGATCA	TTCTCCTGAT	TTATATTCAC	TGGAGCTGGC	AGGTTTGCAT	660
GAAATTGGGA	AGCGTTATGG	GGAAGACTCT	GAACAATTCA	GAGATGCTTC	TAAGATCCTT	720
GTTGACGCTC	TGCAAAAGTT	TGCAGATGAC	ATGTACAGTC	TTTATGGTGG	GAATGCAGTG	780
GTAGAGTTAG	TCACTGTCAA	GTCATTTGAC	ACCTCCCTCA	TTAGGAAGAC	AAGGACTATC	840
CTTGAGGCAA	AACAAGCGAA	GAACCCAGCA	AGTCCCTATA	ACCTTGCATA	TAAGTATAAT	900
TTTGAATATT	CCGTGGTTTT	CAACATGGTA	CTTTGGATAA	TGATCGCCTT	GGCCTTGGCT	960
GTGATTATCA	CCTCTTACAA	TATTTGGAAC	ATGGATCCTG	GATATGATAG	CATCATTTAT	1020
AGGATGACAA	ACCAGAAGAT	TCGAATGGAT				1050

Sequence No.: 32

Sequence length: 627

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Fibrosarcoma

Cell line: HT-1080

Clone name: HP10034

Sequence description

ATGGTGTCTT	CTCCCTGCAC	GCAGGCAAGC	TCACGGACTT	GCTCCCGTAT	CCTGGGACTG	60
AGCCTTGGGA	CTGCAGCCCT	GTTTGCTGCT	GGGGCCAACG	TGGCACTCCT	CCTTCCTAAC	120
TGGGATGTCA	CCTACCTGTT	GAGGGGCCTC	CTTGGCAGGC	ATGCCATGCT	GGGAACTGGG	180
CTCTGGGGAG	GAGGCCTCAT	GGTACTCACT	GCAGCTATCC	TCATCTCCTT	GATGGGCTGG	240
AGATACGGCT	GCTTCAGTAA	GAGTGGGCTC	TGTCGAAGCG	TGCTTACTGC	TCTGTTGTCA	300
GGTGGCCTGG	CTTTACTTGG	AGCCCTGATT	TGCTTTGTCA	CTTCTGGAGT	TGCTCTGAAA	360
GATGGTCCTT	TTTGCATGTT	TGATGTTTCA	TCCTTCAATC	AGACACAAGC	TTGGAAATAT	420
GGTTACCCAT	TCAAAGACCT	GCATAGTAGG	AATTATCTGT	ATGACCGTTC	GCTCTGGAAC	480
TCCGTCTGCC	TGGAGCCCTC	TGCAGCTGTT	GTCTGGCAGC	TGTCCCTCTT	CTCCGCCCTT	540
CTGTGCATCA	GCCTGCTCCA	GCTTCTCCTG	GTGGTCGTTT	ATGTCATCAA	CAGCCTCCTG	600
GGCCTTTTCT	GCAGCCTCTG	CGAGAAG				627

Sequence No.: 33

Sequence length: 489

Sequence type: Nucleic acid

Strandedness: Double



Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Fibrosarcoma

Cell line: HT-1080

Clone name: HP10050

Sequence description

ATGCCGGCTG	GGCTGTTTGG	TTTGAGCGCT	CGCCGTCTTT	TGGCGGCAGC	GGCGACGCGA	60
GGCTCCCCG	CGCCCCGCT	CCGCTGGGAA	TCTAGCTTCT	CCAGGACTGT	GGTCGCCCCG	120
TCCGCTGTGG	CGGGAAAGCG	GCCCCCAGAA	CCGACCACAC	CGTGGCAAGA	GGACCCAGAA	180
CCCGAGGACG	AAAAC TTGTA	TGAGAAGAAC	CCAGACTCCC	ATGGTTATGA	CAAGGACCCC	240
GTTTTGGACG	TCTGGAACAT	GCGACTTGTC	TTCTTCTTTG	GCGTCTCCAT	CATCCTGGTC	300
CTTGGCAGCA	CCTTTGTGGC	CTATCTGCCT	GACTACAGGT	GCACAGGGTG	TCCAAGAGCG	360
TGGGATGGGA	TGAAAGAGTG	GTCCCGCCGC	GAAGCTGAGA	GGCTTGTGAA	ATACCGAGAG	420
GCCAATGGCC	TTCCCATCAT	GGAATCCAAC	TGCTTCGACC	CCAGCAAGAT	CCAGCTGCCA	480
GAGGATGAG						489

Sequence No.: 34

Sequence length: 276

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10071

Sequence description

ATGACGAAAT	TAGCGCAGTG	GCTTTGGGGA	CTAGCGATCC	TGGGCTCCAC	CTGGGTGGCC	60
CTGACCACGG	GAGCCTTGGG	CCTGGAGCTG	CCCTTGTCTT	GCCAGGAAGT	CCTGTGGCCA	120
CTGCCCGCCT	ACTTGCTGGT	GTCCGCCGGC	TGCTATGCCC	TGGGCACTGT	GGGCTATCGT	180
GTGGCCACTT	TTCATGACTG	CGAGGACGCC	GCACGCGAGC	TGCAGAGCCA	GATACAGGAG	240
GCCCGAGCCG	ACTTAGCCCG	CAGGGGGCTG	CGCTTC			276

Sequence No.: 35

Sequence length: 516

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Lymphoma

Cell line: U937

Clone name: HP10076

Sequence description

```
ATGGAATATT TGGCTCATCC CAGTACACTC GGCTTGGCTG TTGGAGTTGC TTGTGGCATG      60
TGCCTGGGCT GGAGCCTTCG AGTATGCTTT GGGATGCTCC CCAAAAGCAA GACGAGCAAG      120
ACACACACAG ATACTGAAAG TGAAGCAAGC ATCTTGGGAG ACACGGGGA GTACAAGATG      180
ATTCTTGTGG TTCGAAATGA CTTAAAGATG GCAAAAGGGA AAGTGGCTGC CCAGTGCTCT      240
CATGCTGCTG TTTCAGCCTA CAAGCAGATT CAAAGAAGAA ATCCTGAAAT GCTCAAACAA      300
TGGGAATACT GTGGCCAGCC CAAGGTGGTG GTCAAAGCTC CTGATGAAGA AACCCTGATT      360
GCATTATTGG CCCATGCAAA AATGCTGGGA CTGACTGTAA GTTTAATTCA AGATGCTGGA      420
CGTACTCAGA TTGCACCAGG CTCTCAAACG GTCCTAGCGA TTGGGCCAGG ACCAGCAGAC      480
CTAATTGACA AAGTCACTGG TCACCTAAAA CTTTAC                                516
```

Sequence No.: 36

Sequence length: 447

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Lymphoma

Cell line: U937

Clone name: HP10085

Sequence description

```
ATGATGACCA AACATAAAAA GTGTTTATA ATTGTTGGTG TTTAATAAC AACTAATATT      60
ATTACTCTGA TAGTTAAACT AACTCGAGAT TCTCAGAGTT TATGCCCTA TGATTGGATT      120
GGTTTCAAA ACAAAATGCTA TTATTTCTCT AAAGAAGAAG GAGATTGGAA TTCAAGTAAA      180
TACAACTGTT CCACTCAACA TGCCGACCTA ACTATAATTG ACAACATAGA AGAAATGAAT      240
TTTCTTAGGC GGTATAAATG CAGTTCTGAT CACTGGATTG GACTGAAGAT GGCAAAAAAT      300
CGAACAGGAC AATGGGTAGA TGGAGCTACA TTTACCAAAT CGTTTGGCAT GAGAGGGAGT      360
GAAGGATGTG CCTACCTCAG CGATGATGGT GCAGCAACAG CTAGATGTTA CACCGAAAGA      420
AAATGGATTT GCAGGAAAAG AATACAC                                447
```

Sequence No.: 37

Sequence length: 564

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10122

Sequence description

ATGAGCACTA	TGTTGCGGGA	CACTCTCCTC	ATCGTTTTTA	TCTCTGTGTG	CACGGCTCTG	60
CTCGCAGAGG	GCATAACCTG	GGTCCTGGTT	TACAGGACAG	ACAAGTACAA	GAGACTGAAG	120
GCAGAAGTGG	AAAAACAGAG	TAAAAAATTG	GAAAAGAAGA	AGGAAACAAT	AACAGAGTCA	180
GCTGGTCGAC	AACAGAAAAA	GAAAATAGAG	AGACAAGAAG	AGAAACTGAA	GAATAACAAC	240
AGAGATCTAT	CAATGGTTTCG	AATGAAATCC	ATGTTTGCTA	TTGGCTTTTG	TTTTACTGCC	300
CTAATGGGAA	TGTTCAATTC	CATATTTGAT	GGTAGAGTGG	TGGCAAAGCT	TCCTTTTACC	360
CCTCTTTCTT	ACATCCAAGG	ACTGTCTCAT	CGAAATCTGC	TGGGAGATGA	CACCACAGAC	420
TGTTCCTTCA	TTTTCCCTGTA	TATTCTCTGT	ACTATGTGGA	TTGACAGAA	CATTCAGAAG	480
ATTCTCGGCC	TTGCCCTTC	ACGAGCCGCC	ACCAAGCAGG	CAGGTGGATT	TCTTGGCCCA	540
CCACCTCCTT	CTGGGAAGTT	CTCT				564

Sequence No.: 38

Sequence length: 645

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Lymphoma

Cell line: U937

Clone name: HP10136

Sequence description

ATGGTGTTC	TAACAATGAT	CGCCCGAGTG	GCGGACGGGC	TCCCGCTGGC	CGCCTCGATG	60
CAGGAGGACG	AACAGTCTGG	CCGGGACCTT	CAACAGTATC	AGAGTCAGGC	TAAGCAATC	120
TTTCGAAAGT	TGAATGAACA	GTCCCCTACC	AGATGTACCT	TGGAAGCAGG	AGCCATGACT	180
TTTCACTACA	TTATTGAGCA	GGGGGTGTGT	TATTTGTTT	TATGTGAAGC	TGCCTTCCCT	240
AAGAAGTTGG	CTTTTGCCTA	CCTAGAAGAT	TTGCACTCAG	AATTTGATGA	ACAGCATGGA	300
AAGAAGGTGC	CCACTGTGTC	CCGACCCTAT	TCCTTTATTG	AATTTGATAC	TTTCATTGAG	360
AAAACCAAGA	AGCTCTACAT	TGACAGTCGT	GCTCGAAGAA	ATCTAGGCTC	CATCAACACT	420
GAATTGCAAG	ATGTGCAGAG	GATCATGGTG	GCCAATATTG	AAGAAGTGTT	ACAACGAGGA	480
GAAGCACTCT	CAGCATTGGA	TTCAAAGGCT	AACAATTTGT	CCAGTCTGTC	CAAGAAATAC	540

CGCCAGGATG CGAAGTACTT GAACATGCGT TCCAATTATG CCAAACCTGC AGCAGTAGCT 600  
GTATTTTTC A TCATGTTAAT AGTGTATGTC CGATTCTGGT GGCTG 645

Sequence No.: 39  
Sequence length: 336  
Sequence type: Nucleic acid  
Strandedness: Double  
Topology: Linear  
Sequence kind: cDNA to mRNA  
Original source:  
Organism species: *Homo sapiens*  
Cell kind: Stomach cancer  
Clone name: HP10175  
Sequence description

ATGCAGGACA CTGGCTCAGT AGTGCCTTTG CATTGGTTTG GCTTTGGCTA CGCAGCACTG 60  
GTTGCTTCTG GTGGGATCAT TGGCTATGTA AAAGCAGGCA GCGTGCCGTC CCTGGCTGCA 120  
GGGCTGCTCT TTGGCAGTCT AGCCGGCCTG GGTGCTTACC AGCTGTCTCA GGATCCAAGG 180  
AACGTTTGGG TTTTCCTAGC TACATCTGGT ACCTTGGCTG GCATTATGGG AATGAGGTTC 240  
TACCACTCTG GAAAATTTCAT GCCTGCAGGT TTAATTGCAG GTGCCAGTTT GCTGATGGTC 300  
GCCAAAGTTG GAGTTAGTAT GTTCAACAGA CCCCAT 336

Sequence No.: 40  
Sequence length: 342  
Sequence type: Nucleic acid  
Strandedness: Double  
Topology: Linear  
Sequence kind: cDNA to mRNA  
Original source:  
Organism species: *Homo sapiens*  
Cell kind: Epidermoid carcinoma  
Cell line: KB  
Clone name: HP10179  
Sequence description

ATGGAGAAGC CCCTCTTCCC ATTAGTGCCT TTGCATTGGT TTGGCTTTGG CTACACAGCA 60  
CTGGTTGTTT CTGGTGGGAT CGTTGGCTAT GTAAAAACAG GCAGCGTGCC GTCCCTGGCT 120  
GCAGGGCTGC TCTTCGGCAG TCTAGCCGGC CTGGGTGCTT ACCAGCTGTA TCAGGATCCA 180  
AGGAACGTTT GGGGTTTCCT AGCCGCTACA TCTGTACTT TTGTTGCTGT TATGGGAATG 240  
AGATCCTACT ACTATGGAAA ATTCATGCCT GTAGGTTTAA TTGCAGGTGC CAGTTTGCTG 300  
ATGGCCGCCA AAGTTGGAGT TCGTATGTTG ATGACATCTG AT 342

Sequence No.: 41  
Sequence length: 981  
Sequence type: Nucleic acid  
Strandedness: Double  
Topology: Linear  
Sequence kind: cDNA to mRNA  
Original source:  
Organism species: *Homo sapiens*  
Cell kind: Fibrosarcoma  
Cell line: HT-1080  
Clone name: HP10196  
Sequence description

ATGGCGGCGG	CGGCGGCGGC	GGCTGCAGCT	ACGAACGGGA	CCGGAGGAAG	CAGCGGGATG	60
GAGGTGGATG	CAGCAGTAGT	CCCCAGCGTG	ATGGCCTGCG	GAGTGACTGG	GAGTGTTC	120
GTCGCTCTCC	ATCCCTTGT	CATTCTCAAC	ATCTCAGACC	ACTGGATCCG	CATGCGCTCC	180
CAGGAGGGGC	GGCCTGTGCA	GGTGATTGGG	GCTCTGATTG	GCAAGCAGGA	GGGCCGAAAT	240
ATCGAGGTGA	TGAACTCCTT	TGAGCTGCTG	TCCCACACCG	TGGAAGAGAA	GATTATCATT	300
GACAAGGAAT	ATTATTACAC	CAAGGAGGAG	CAGTTTAAAC	AGGTGTTCAA	GGAGCTGGAG	360
TTTCTGGGTT	GGTATACCAC	AGGGGGGCCA	CCTGACCCCT	CGGACATCCA	CGTCCATAAG	420
CAGGTGTGTG	AGATCATCGA	GAGCCCCCTC	TTTCTGAAGT	TGAACCCTAT	GACCAAGCAC	480
ACAGATCTTC	CTGTCAGCGT	TTTTGAGTCT	GTCATTGATA	TAATCAATGG	AGAGGCCACA	540
ATGCTGTTTG	CTGAGCTGAC	CTACACTCTG	GCCACAGAGG	AAGCGGAACG	CATTGGTGTA	600
GACCACGTAG	CCCGAATGAC	AGCAACAGGC	AGTGGAGAGA	ACTCCACTGT	GGCTGAACAC	660
CTGATAGCAC	AGCACAGCGC	CATCAAGATG	CTGCACAGCC	GCGTCAAGCT	CATCTTGGAG	720
TACGTCAAGG	CCTCTGAAGC	GGGAGAGGTC	CCCTTTAATC	ATGAGATCCT	GCGGGAGGCC	780
TATGCTCTGT	GTCAGTGTCT	CCCGGTGCTC	AGCACAGACA	AGTTCAAGAC	AGATTTTTAT	840
GATCAATGCA	ACGACGTGGG	GCTCATGGCC	TACCTCGGCA	CCATCACCAA	AACGTGCAAC	900
ACCATGAACC	AGTTTGTGAA	CAAGTTCAAT	GTCCTCTACG	ACCGACAAGG	CATCGGCAGG	960
AGAATGCGCG	GGCTCTTTTT	C				981

Sequence No.: 42  
Sequence length: 1119  
Sequence type: Nucleic acid  
Strandedness: Double  
Topology: Linear  
Sequence kind: cDNA to mRNA  
Original source:  
Organism species: *Homo sapiens*  
Cell kind: Fibrosarcoma  
Cell line: HT-1080  
Clone name: HP10235  
Sequence description

ATGACCCTAT GTGCCATGCT GCCCCTGCTG TTATTACCT ACCTCAACTC CTTCTGCAT	60
CAGAGGATCC CCCAGTCCGT ACGGATCCTG GGCAGCCTGG TGGCCATCCT GCTGGTGT	120
CTGATCACTG CCATCCTGGT GAAGGTGCAG CTGGATGCTC TGCCCTTCTT TGTCATCACC	180
ATGATCAAGA TCGTGCTCAT TAATTCATTT GGTGCCATCC TGCAGGGCAG CCTGTTTGGT	240
CTGGCTGGCC TTCTGCCTGC CAGCTACACG GCCCCATCA TGAGTGGCCA GGGCCTAGCA	300
GGCTTCTTTG CCTCCGTGGC CATGATCTGC GCTATTGCCA GTGGCTCGGA GCTATCAGAA	360
AGTGCCTTCG GCTACTTTAT CACAGCCTGT GCTGTTATCA TTTTGACCAT CATCTGTTAC	420
CTGGGCCTGC CCCGCCCTGA ATTCTACCGC TACTACCAGC AGCTCAAGCT TGAAGGACCC	480
GGGGAGCAGG AGACCAAGTT GGACCTCATT AGCAAAGGAG AGGAGCCAAG AGCAGGCAAA	540
GAGGAATCTG GAGTTTCAGT CTCCAACTCT CAGCCCACCA ATGAAAGCCA CTCTATCAAA	600
GCCATCCTGA AAAATATCTC AGTCCTGGCT TTCTCTGTCT GCTTCATCTT CACTATCACC	660
ATTGGGATGT TTCCAGCCGT GACTGTTGAG GTCAAGTCCA GCATCGCAGG CAGCAGCACC	720
TGGGAACGTT ACTTCATTCC TGTGTCTGT TTCTTGACTT TCAATATCTT TGACTGGTTG	780
GGCCGGAGCC TCACAGCTGT ATTCATGTGG CCTGGGAAGG ACAGCCGCTG GCTGCCAAGC	840
CTGGTGCTGG CCCGGCTGGT GTTTGTGCCA CTGCTGCTGC TGTGCAACAT TAAGCCCCGC	900
CGCTACCTGA CTGTGGTCTT CGAGCAGCAT GCCTGGTTCA TCTTCTTCAT GGCTGCCTTT	960
GCCTTCTCCA ACGGCTACCT CGCCAGCCTC TGCATGTGCT TCGGGCCCAA GAAAGTGAAG	1020
CCAGCTGAGG CAGAGACCGC AGGAGCCATC ATGGCCTTCT TCCTGTGTCT GGGTCTGGCA	1080
CTGGGGGCTG TTTTCTCCTT CCTGTTCCGG GCAATTGTG	1119

Sequence No.: 43

Sequence length: 549

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10297

Sequence description

ATGAAGCTCT TATCTTTGGT GGCTGTGGTC GGGTGTTCG TGGTGCCCCC AGCTGAAGCC	60
AACAAGAGTT CTGAAGATAT CCGGTGCAAA TGCATCTGTC CACCTTATAG AAACATCAGT	120
GGGCACATTT ACAACCAGAA TGTATCCAG AAGGACTGCA ACTGCCTGCA CGTGGTGGAG	180
CCCATGCCAG TGCCTGGCCA TGACGTGGAG GCCTACTGCC TGCTGTGCGA GTGCAGGTAC	240
GAGGAGCGCA GCACCACCAC CATCAAGGTC ATCATTGTCA TCTACCTGTC CGTGGTGGGT	300
GCCCTGTTGC TCTACATGGC CTTCTGATG CTGGTGGACC CTCTGATCCG AAAGCCGGAT	360
GCATACACTG AGCAACTGCA CAATGAGGAG GAGAATGAGG ATGCTCGCTC TATGGCAGCA	420
GCTGCTGCAT CCCTCGGGGG ACCCCGAGCA AACACAGTCC TGGAGCGTGT GGAAGGTGCC	480
CAGCAGCGGT GGAAGCTGCA GGTGCAGGAG CAGCGGAAGA CAGTCTTCGA TCGGCACAAG	540
ATGCTCAGC	549

Sequence No.: 44  
Sequence length: 348  
Sequence type: Nucleic acid  
Strandedness: Double  
Topology: Linear  
Sequence kind: cDNA to mRNA  
Original source:  
Organism species: *Homo sapiens*  
Cell kind: Stomach cancer  
Clone name: HP10299  
Sequence description

ATGGCCAGTA	CAGTGGTAGC	AGTTGGACTG	ACCATTGCTG	CTGCAGGATT	TGCAGGCCGT	60
TACGTTTTGC	AAGCCATGAA	GCATATGGAG	CCTCAAGTAA	AACAAGTTT	TCAAAGCCTA	120
CCAAAATCTG	CCTTCAGTGG	TGGCTATTAT	AGAGGTGGGT	TTGAACCCAA	AATGACAAAA	180
CGGGAAGCA	GCATTAATAC	TAGGTGTAAG	CCCTACTGCC	AATAAAGGGA	AAATAAGAGA	240
GCTCATCGAC	GAATTATGCT	TTTAAATCAT	CCTGACAAAG	GAGGATCTCC	TTATATAGCA	300
GCCAAAATCA	ATGAAGCTAA	AGATTTACTA	GAAGGTCAAG	CTAAAAAA		348

Sequence No.: 45  
Sequence length: 456  
Sequence type: Nucleic acid  
Strandedness: Double  
Topology: Linear  
Sequence kind: cDNA to mRNA  
Original source:  
Organism species: *Homo sapiens*  
Cell kind: Epidermoid carcinoma  
Cell line: KB  
Clone name: HP10301  
Sequence description

ATGGCTGTCC	TCTCTAAGGA	ATATGGTTTT	GTGCTTCTAA	CTGGTGCTGC	CAGCTTTATA	60
ATGGTGCC	ACCTAGCCAT	CAATGTTTCC	AAGGCCCGCA	AGAAGTACAA	AGTGGAGTAT	120
CCTATCATGT	ACAGCACGGA	CCCTGAAAAT	GGGCACATCT	TCAACTGCAT	TCAGCGAGCC	180
CACCAGAACA	CGTTGGAAGT	GTATCCTCCC	TTCTTATTTT	TTCTAGCTGT	TGGAGGTGTT	240
TACCACCCGC	GTATAGCTTC	TGGCCTGGGC	TTGGCCTGGA	TTGTTGGACG	AGTTCTTTAT	300
GCTTATGGCT	ATTACACGGG	AGAACCCAGC	AAGCGTAGTC	GAGGAGCCCT	GGGGTCCATC	360
GCCCTCCTGG	GCTTGGTGGG	CACAACTGTG	TGCTCTGCTT	TCCAGCATCT	TGGTTGGGTT	420
AAAAGTGGCT	TGGGCAGTGG	ACCCAAATGC	TGCCAT			456

Sequence No.: 46  
Sequence length: 1677  
Sequence type: Nucleic acid  
Strandedness: Double  
Topology: Linear  
Sequence kind: cDNA to mRNA  
Original source:  
Organism species: *Homo sapiens*  
Cell kind: Liver  
Clone name: HP10302  
Sequence description

ATGGCCCCCA	CGCTGCAACA	GGCGTACCGG	AGGCGCTGGT	GGATGGCCTG	CACGGCTGTG	60
CTGGAGAACC	TCTTCTTCTC	TGCTGTACTC	CTGGGCTGGG	GCTCCCTGTT	GATCATTCTG	120
AAGAACGAGG	GCTTCTATTG	CAGCACGTGC	CCAGCTGAGA	GCAGCACCAA	CACCACCCAG	180
GATGAGCAGC	GCAGGTGGCC	AGCCTGTGAC	CAGCAGGACG	AGATGCTCAA	CCTGGGCTTC	240
ACCATTTGTT	CCTTCGTGCT	CAGCGCCACC	ACCCTGCCAC	TGGGGATCCT	CATGGACCGC	300
TTTGGCCCCC	GACCCGTGCG	GCTGGTTGGC	AGTGCCTGCT	TCATGCGGTC	CTGCACCCTC	360
ATGGCCCTGG	CCTCCCGGGA	CGTGGAAGCT	CTGTCTCCGT	TGATATTCTT	GGCGCTGTCC	420
CTGAATGGCT	TTGGTGGCAT	CTGCCTAACG	TTCACCTTAC	TCAGCTGCC	CAACATGTTT	480
GGGAACCTGC	GCTCCACGTT	AATGGCCCTC	ATGATTGGCT	CTTACGCCTC	TTCTGCCATT	540
ACGTTCCCAG	GAATCAAGCT	GATCTACGAT	GCCGGTGTGG	CCTTCGTGGT	CATCATGTTT	600
ACCTGGTCTG	GCCTGGCCTG	CCTTATCTTT	CTGAAGTACA	CCCTCAACTG	GCCCATCGAA	660
GCCTTTCTCT	CCCCTGAGGA	AGTCAATTAC	ACGAAGAAGA	TCAAGCTGAG	TGGGCTGGCC	720
CTGGACCACA	AGGTGACAGG	TGACCTCTTC	TACACCCATG	TGACCACCAT	GGGCCAGAGG	780
CTCAGCCAGA	AGGCCCCAG	CCTGGAGGAC	GGTTCGGATG	CCTTCATGTC	ACCCAGGAT	840
GTTCCGGGCA	CCTCAGAAAA	CCTTCCTGAG	AGGTCTGTCC	CCTTACGCAA	GAGCCTCTGC	900
TCCCCCACTT	TCCTGTGGAG	CCTCCTCACC	ATGGGCATGA	CCCAGCTGCG	GATCATCTTC	960
TACATGGCTG	CTGTGAACAA	GATGCTGGAG	TACCTTGTGA	CTGGTGGCCA	GGAGCATGAG	1020
ACAAATGAAC	AGCAACAAAA	GGTGGCAGAG	ACAGTTGGGT	TCTACTCCTC	CGTCTTCGGG	1080
GCCATGCAGC	TGTTGTGCCT	TCTCACCTGC	CCCCTCATTG	GCTACATCAT	GGACTGGCGG	1140
ATCAAGGACT	GCGTGGACGC	CCCAACTCAG	GGCACTGTCC	TCGGAGATGC	CAGGGACGGG	1200
GTTGCTACCA	AATCCATCAG	ACCACGCTAC	TGCAAGATCC	AAAAGCTCAC	CAATGCCATC	1260
AGTGCCTTCA	CCCTGACCAA	CCTGCTGCTT	GTGGGTTTTG	GCATCACCTG	TCTCATCAAC	1320
AACTTACACC	TCCAGTTTGT	GACCTTTGTC	CTGCACACCA	TTGTTCCAGG	TTTCTTCCAC	1380
TCAGCCTGTG	GGAGTCTCTA	TGCTGCAGTG	TTCCCATCCA	ACCACTTTGG	GACGCTGACA	1440
GGCCTGCAGT	CCCTCATCAG	TGCTGTGTTT	GCCTTGCTTC	AGCAGCCACT	TTTCATGGCG	1500
ATGGTGGGAC	CCCTGAAAGG	AGAGCCCTTC	TGGGTGAATC	TGGGCCTCCT	GCTATTCTCA	1560
CTCCTGGGAT	TCCTGTTGCC	TTCCTACCTC	TTCTATTACC	GTGCCCCGCT	CCAGCAGGAG	1620
TACGCCGCCA	ATGGGATGGG	CCCACTGAAG	GTGCTTAGCG	GCTCTGAGGT	GACCGCA	1677

Sequence No.: 47  
Sequence length: 990



Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Osteosarcoma

Cell line: U-2 OS

Clone name: HP10304

Sequence description

```
ATGGAGGGCG CTCCACCGGG GTCGCTCGCC CTCGGGCTCC TGCTGTTTCT GGCCTACCC 60
GCCTCCGGCT GGCTGACGAC GGGCGCCCCC GAGCCGCCGC CGCTGTCCGG AGCCCCACAG 120
GACGGCATCA GAATTAATGT AACTACACTG AAAGATGATG GGGACATATC TAAACAGCAG 180
GTTGTTCTTA ACATAACCTA TGAGAGTGGA CAGGTGTATG TAAATGACTT ACCTGTAAAT 240
AGTGGTGTA CCCGAATAAG CTGTCAGACT TTGATAGTGA AGAATGAAAA TCTTGAAAAAT 300
TTGGAGGAAA AAGAATATTT TGAATTGTC AGTGTAAAGG TTTTAGTTCA TGAGTGGCCT 360
ATGACATCTG GTTCCAGTTT GCAACTAATT GTCATTCAAG AAGAGGTAGT AGAGATTGAT 420
GGAAAACAAG TTCAGCAAAA GGATGTCAGT GAAATTGATA TTTTAGTTAA GAACCGGGGA 480
GTAATCAGAC ATTCAAATA TACCCTCCCT TTGGAAGAAA GCATGCTCTA CTCTATTTCT 540
CGAGACAGTG ACATTTTATT TACCCTTCCT AACCTCTCCA AAAAAGAAAAG TGTTAGTTCA 600
CTGCAAACCA CTAGCCAGTA TCTTATCAGG AATGTGGAAG CCACTGTAGA TGAAGATGTT 660
TTACCTGGCA AGTTACCTGA AACTCCTCTC AGAGCAGAGC CGCCATCTTC ATATAAGGTA 720
ATGTGTCACT GGATGGAAAA GTTTAGAAAA GATCTGTGTA GGTTCCTGGAG CAACGTTTTT 780
CCAGTATTCT TTCAGTTTTT GAACATCATG GTGGTTGGAA TTACAGGAGC AGCTGTGGTA 840
ATAACCATCT TAAAGGTGTT TTTCCAGTT TCTGAATACA AAGGAATTCT TCAGTTGGAT 900
AAAGTGGACG TCATACCTGT GACAGCTATC AACTTATATC CAGATGGTCC AGAGAAAAGA 960
GCTGAAAACC TTGAAGATAA AACATGTATT 990
```

Sequence No.: 48

Sequence length: 324

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Osteosarcoma

Cell line: U-2 OS

Clone name: HP10305

Sequence description

```
ATGAGTCTGA CTTCCAGTTC CAGCGTACGA GTTGAATGGA TCGCAGCAGT TACCATTGCT 60
```

GCTGGGACAG CTGCAATTGG TTATCTAGCT TACAAAAGAT TTTATGTTAA AGATCATCGA	120
AATAAAGCTA TGATAAACCT TCACATCCAG AAAGACAACC CCAAGATAGT ACATGCTTTT	180
GACATGGAGG ATTTGGGAGA TAAAGCTGTG TACTGCCGTT GTTGGAGGTC CAAAAAGTTC	240
CCATTCTGTG ATGGGGCTCA CACAAAACAT AACGAAGAGA CTGGAGACAA TGTGGGCCCT	300
CTGATCATCA AGAAAAAAGA AACT	324

Sequence No.: 49

Sequence length: 303

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Osteosarcoma

Cell line: U-2 OS

Clone name: HP10306

Sequence description

ATGAACCTGG AGCGAGTGTC CAATGAGGAG AAATTGAACC TGTGCCGGAA GTACTACCTG	60
GGGGGGTTTG CTTTCCTGCC TTTTCTCTGG TTGGTCAACA TCTTCTGGTT CTTCCGAGAG	120
GCCTTCCTTG TCCCAGCCTA CACAGAACAG AGCCAAATCA AAGGCTATGT CTGGCGCTCA	180
GCTGTGGGCT TCCTCTTCTG GGTGATAGTG CTCACCTCCT GGATCACCAT CTTCCAGATC	240
TACCGGCCCC GCTGGGGTGC CCTTGGGGAC TACCTCTCCT TCACCATAACC CCTGGGCACC	300
CCC	303

Sequence No.: 50

Sequence length: 1116

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP10328

Sequence description

ATGAAGTATC TCCGGCACCG GCGGCCCAAT GCCACCCTCA TTCTGGCCAT CGGCGCTTTC	60
ACCCTCCTCC TCTTCAGTCT GCTAGTGTC AACCACCT GCAAGGTCCA GGAGCAGCCA	120
CCGGCGATCC CCGAGGCCCT GGCCTGGCCC ACTCCACCCA CCCGCCCAGC CCCGGCCCCG	180

TGCCATGCCA	ACACCTCTAT	GGTCACCCAC	CCGGA CTTCG	CCACGCAGCC	GCAGCACGTT	240
CAGAACTTCC	TCCTGTACAG	ACACTGCCGC	CACTTTCCCC	TGCTGCAGGA	CGTGCCCCCC	300
TCTAAGTGCG	CGCAGCCGGT	CTTCCTGCTG	CTGGTGATCA	AGTCCTCCCC	TAGCAACTAT	360
GTGCGCCGCG	AGCTGCTGCG	GCGCACGTGG	GGCCGCGAGC	GCAAGGTACG	GGGTTTGCAG	420
CTGCGCCTCC	TCTTCCTGGT	GGGCACAGCC	TCCAACCCGC	ACGAGGCCCG	CAAGGTCAAC	480
CGGCTGCTGG	AGCTGGAGGC	ACAGACTCAC	GGAGACATCC	TGCAGTGGGA	CTTCCACGAC	540
TCCTTCTTCA	ACCTCACGCT	CAAGCAGGTC	CTGTTCTTAC	AGTGGCAGGA	GACAAGGTGC	600
GCCAACGCCA	GCTTCGTGCT	CAACGGGGAT	GATGACGTCT	TTGCACACAC	AGACAACATG	660
GTCTTCTACC	TGCAGGACCA	TGACCCTGGC	CGCCACCTCT	TCGTGGGGCA	ACTGATCCAA	720
AACGTGGGCC	CCATCCGGGC	TTTTTGAGC	AAGTACTATG	TGCCAGAGGT	GGTGACTCAG	780
AATGAGCGGT	ACCCACCCTA	TTGTGGGGGT	GGTGGCTTCT	TGCTGTCCCG	CTTCACGGCC	840
GCTGCCCTGC	GCCGTGCTGC	CCATGTCTTG	GACATCTTCC	CCATTGATGA	TGTCTTCCTG	900
GGTATGTGTC	TGGAGCTTGA	GGGACTGAAG	CCTGCCTCCC	ACAGCGGCAT	CCGCACGTCT	960
GGCGTGCGGG	CTCCATCGCA	ACACCTGTCC	TCCTTTGACC	CCTGCTTCTA	CCGAGACCTG	1020
CTGCTGGTGC	ACCGCTTCCT	ACCTTATGAG	ATGCTGCTCA	TGTGGGATGC	GCTGAACCAG	1080
CCCAACCTCA	CCTGCGGCAA	TCAGACACAG	ATCTAC			1116

Sequence No.: 51

Sequence length: 986

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Fibrosarcoma

Cell line: HT-1080

Clone name: HP00442

Sequence characteristics

Code representing characteristics: CDS

Existence site: 82.. 699

Characterization method: E

Sequence description

AGACTGCGGG	ACGGACGGTG	GACGCTGGGA	CGCGTTTGTA	GCTCCGGCCC	CGCCGTTCGG	60
ACCCCCGCCG	CCGTGCGCCG	C ATG ACG GGG	CTA GCA CTG	CTC TAC TCC	GGG	111
Met Thr Gly Leu Ala Leu Leu Tyr Ser Gly						
1 5 10						
GTC TTC GTG	GCC TTC TGG	GCC TGC GCG	CTG GCC GTG	GGA GTC TGC	TAC	159
Val Phe Val	Ala Phe Trp	Ala Cys Ala	Leu Ala Val	Gly Val Cys	Tyr	
15 20 25						
ACC ATT TTT	GAT TTG GGC	TTC CGC TTT	GAT GTG GCA	TGG TTC CTG	ACG	207
Thr Ile Phe	Asp Leu Gly	Phe Arg Phe	Asp Val Ala	Trp Phe Leu	Thr	

30	35	40	
GAG ACT TCG CCC TTC ATG TGG TCC AAC CTG GGC ATT GGC CTA GCT ATC			255
Glu Thr Ser Pro Phe Met Trp Ser Asn Leu Gly Ile Gly Leu Ala Ile			
45	50	55	
TCC CTG TCT GTG GTT GGG GCA GCC TGG GGC ATC TAT ATT ACC GGC TCC			303
Ser Leu Ser Val Val Gly Ala Ala Trp Gly Ile Tyr Ile Thr Gly Ser			
60	65	70	
TCC ATC ATT GGT GGA GGA GTG AAG GCC CCC AGG ATC AAG ACC AAG AAC			351
Ser Ile Ile Gly Gly Gly Val Lys Ala Pro Arg Ile Lys Thr Lys Asn			
75	80	85	90
CTG GTC AGC ATC ATC TTC TGT GAG GCT GTG GCC ATC TAC GGC ATC ATC			399
Leu Val Ser Ile Ile Phe Cys Glu Ala Val Ala Ile Tyr Gly Ile Ile			
95	100	105	
ATG GCA ATT GTC ATT AGC AAC ATG GCT GAG CCT TTC AGT GCC ACA GAC			447
Met Ala Ile Val Ile Ser Asn Met Ala Glu Pro Phe Ser Ala Thr Asp			
110	115	120	
CCC AAG GCC ATC GGC CAT CGG AAC TAC CAT GCA GGC TAC TCC ATG TTT			495
Pro Lys Ala Ile Gly His Arg Asn Tyr His Ala Gly Tyr Ser Met Phe			
125	130	135	
GGG GCT GGC CTC ACC GTA GGC CTG TCT AAC CTC TTC TGT GGA GTC TGC			543
Gly Ala Gly Leu Thr Val Gly Leu Ser Asn Leu Phe Cys Gly Val Cys			
140	145	150	
GTG GGC ATC GTG GGC AGT GGG GCT GCC CTG GCC GAT GCT CAG AAC CCC			591
Val Gly Ile Val Gly Ser Gly Ala Ala Leu Ala Asp Ala Gln Asn Pro			
155	160	165	170
AGC CTC TTT GTA AAG ATT CTC ATC GTG GAG ATC TTT GGC AGC GCC ATT			639
Ser Leu Phe Val Lys Ile Leu Ile Val Glu Ile Phe Gly Ser Ala Ile			
175	180	185	
GGC CTC TTT GGG GTC ATC GTC GCA ATT CTT CAG ACC TCC AGA GTG AAG			687
Gly Leu Phe Gly Val Ile Val Ala Ile Leu Gln Thr Ser Arg Val Lys			
190	195	200	
ATG GGT GAC TAGATGATAT GTGTGGGTGG GGCCGTGCCT CACT			730
Met Gly Asp			
205			
TTTATTTATT GCTGGTTTTTCTCTGGGACAGC TGGAGCTGTG TCCCTTAGCC TTTCAGAGGC			790
TTGGTGTTCA GGGCCCTCCC TGGACTCCCC TCTTGCTGCG TGTTGATTTC GAGGCACTGC			850
AGTCCAGGCC GAGTCCTCAG TCGGGGAGC AGGCTGCTGC TGCTGACTCT GTGCAGCTGC			910
GCACCTGTGT CCCCCACCTC CACCCTCAAC CCATCTTCCT AGTGTTTGTG AAATAAACTT			970
GGTATTTGTC TGGGTC			986

Sequence No.: 52

Sequence length: 1824

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Leukocyte

Clone name: HP00804

Sequence characteristics

Code representing characteristics: CDS

Existence site: 133.. 1248

Characterization method: E

Sequence description

```

GGCCCAGCTG AGCGGCCGCC GAGCGGGTGC GGGTGC GGGC GCATCGGCCA TCACCGCGCG      60
GCCGCGCAGC GGACACCGTG CGTACCGGCC TCGCGCGCCC GGCCACCGGG GCGGACCGCG      120
GAACCCGAGG CC ATG TCC CAT GAA AAG AGT TTT TTG GTG TCT GGG GAC AAC      171
      Met Ser His Glu Lys Ser Phe Leu Val Ser Gly Asp Asn
          1             5             10
TAT CCT CCC CCC AAC CCT GGA TAT CCG GGG GGG CCC CAG CCA CCC ATG      219
Tyr Pro Pro Pro Asn Pro Gly Tyr Pro Gly Gly Pro Gln Pro Pro Met
      15             20             25
CCC CCC TAT GCT CAG CCT CCC TAC CCT GGG GCC CCT TAC CCA CAG CCC      267
Pro Pro Tyr Ala Gln Pro Pro Tyr Pro Gly Ala Pro Tyr Pro Gln Pro
      30             35             40             45
CCT TTC CAG CCC TCC CCC TAC GGT CAG CCA GGG TAC CCC CAT GGC CCC      315
Pro Phe Gln Pro Ser Pro Tyr Gly Gln Pro Gly Tyr Pro His Gly Pro
          50             55             60
AGC CCC TAC CCC CAA GGG GGC TAC CCA CAG GGT CCC TAC CCC CAA GGG      363
Ser Pro Tyr Pro Gln Gly Gly Tyr Pro Gln Gly Pro Tyr Pro Gln Gly
          65             70             75
GGC TAC CCA CAG GGC CCC TAC CCA CAA GAG GGC TAC CCA CAG GGC CCC      411
Gly Tyr Pro Gln Gly Pro Tyr Pro Gln Glu Gly Tyr Pro Gln Gly Pro
          80             85             90
TAC CCC CAA GGG GGC TAC CCC CAG GGG CCA TAT CCC CAG AGC CCC TTC      459
Tyr Pro Gln Gly Gly Tyr Pro Gln Gly Pro Tyr Pro Gln Ser Pro Phe
          95             100             105
CCC CCC AAC CCC TAT GGA CAG CCA CAG GTC TTC CCA GGA CAA GAC CCT      507
Pro Pro Asn Pro Tyr Gly Gln Pro Gln Val Phe Pro Gly Gln Asp Pro
      110             115             120             125
GAC TCA CCC CAG CAT GGA AAC TAC CAG GAG GAG GGT CCC CCA TCC TAC      555
Asp Ser Pro Gln His Gly Asn Tyr Gln Glu Glu Gly Pro Pro Ser Tyr
          130             135             140
TAT GAC AAC CAG GAC TTC CCT GCC ACC AAC TGG GAT GAC AAG AGC ATC      603
Tyr Asp Asn Gln Asp Phe Pro Ala Thr Asn Trp Asp Asp Lys Ser Ile

```

136

145	150	155	
CGA CAG GCC TTC ATC CGC AAG GTG TTC CTA GTG CTG ACC TTG CAG CTG			651
Arg Gln Ala Phe Ile Arg Lys Val Phe Leu Val Leu Thr Leu Gln Leu			
160	165	170	
TCG GTG ACC CTG TCC ACG GTG TCT GTG TTC ACT TTT GTT GCG GAG GTG			699
Ser Val Thr Leu Ser Thr Val Ser Val Phe Thr Phe Val Ala Glu Val			
175	180	185	
AAG GGC TTT GTC CGG GAG AAT GTC TGG ACC TAC TAT GTC TCC TAT GCT			747
Lys Gly Phe Val Arg Glu Asn Val Trp Thr Tyr Tyr Val Ser Tyr Ala			
190	195	200	205
GTC TTC TTC ATC TCT CTC ATC GTC CTC AGC TGT TGT GGG GAC TTC CGG			795
Val Phe Phe Ile Ser Leu Ile Val Leu Ser Cys Cys Gly Asp Phe Arg			
210	215	220	
CGA AAG CAC CCC TGG AAC CTT GTT GCA CTG TCG GTC CTG ACC GCC AGC			843
Arg Lys His Pro Trp Asn Leu Val Ala Leu Ser Val Leu Thr Ala Ser			
225	230	235	
CTG TCG TAC ATG GTG GGG ATG ATC GCC AGC TTC TAC AAC ACC GAG GCA			891
Leu Ser Tyr Met Val Gly Met Ile Ala Ser Phe Tyr Asn Thr Glu Ala			
240	245	250	
GTC ATC ATG GCC GTG GGC ATC ACC ACA GCC GTC TGC TTC ACC GTC GTC			939
Val Ile Met Ala Val Gly Ile Thr Thr Ala Val Cys Phe Thr Val Val			
255	260	265	
ATC TTC TCC ATG CAG ACC CGC TAC GAC TTC ACC TCA TGC ATG GGC GTG			987
Ile Phe Ser Met Gln Thr Arg Tyr Asp Phe Thr Ser Cys Met Gly Val			
270	275	280	285
CTC CTG GTG AGC ATG GTG GTG CTC TTC ATC TTC GCC ATT CTC TGC ATC			1035
Leu Leu Val Ser Met Val Val Leu Phe Ile Phe Ala Ile Leu Cys Ile			
290	295	300	
TTC ATC CGG AAC CGC ATC CTG GAG ATC GTG TAC GCC TCA CTG GGC GCT			1083
Phe Ile Arg Asn Arg Ile Leu Glu Ile Val Tyr Ala Ser Leu Gly Ala			
305	310	315	
CTG CTC TTC ACC TGC TTC CTC GCA GTG GAC ACC CAG CTG CTG CTG GGG			1131
Leu Leu Phe Thr Cys Phe Leu Ala Val Asp Thr Gln Leu Leu Leu Gly			
320	325	330	
AAC AAG CAG CTG TCC CTG AGC CCA GAA GAG TAT GTG TTT GCT GCG CTG			1179
Asn Lys Gln Leu Ser Leu Ser Pro Glu Glu Tyr Val Phe Ala Ala Leu			
335	340	345	
AAC CTG TAC ACA GAC ATC ATC AAC ATC TTC CTG TAC ATC CTC ACC ATC			1227
Asn Leu Tyr Thr Asp Ile Ile Asn Ile Phe Leu Tyr Ile Leu Thr Ile			
350	355	360	365
ATT GGC CGC GCC AAG GAG TAGCCGAGCT CCAGCTCGCT GTGCC			1270
Ile Gly Arg Ala Lys Glu			
370			
CGCTCAGGTG GCACGGCTGG CCTGGACCCT GCCCCTGGCA CGGCAGTGCC AGCTGTACTT			1330

```

CCCCCTCTCTC TTGTCCCCAG GCACAGCCTA GGGAAAAGGA TGCCTCTCTC CAACCCTCCT 1390
GTATGTACAC TGCAGATACT TCCATTTGGA CCCGCTGTGG CCACAGCATG GCCCCTTTAG 1450
TCCTCCCGCC CCCGCCAAGG GGCACCAAGG CCACGTTTCC GTGCCACCTC CTGTCTACTC 1510
ATTGTTGCAT GAGCCCTGTC TGCCAGCCCA CCCAGGGAC TGGGGGCAGC ACCAGGTCCC 1570
GGGGAGAGGG ATTGAGCCAA GAGGTGAGGG TGCACGTCTT CCCTCCTGTC CCAGCTCCCC 1630
AGCCTGGCGT AGAGCACCCC TCCCCTCCCC CCCACCCCC TGGAGTGCTG CCCTCTGGGG 1690
ACATGCGGAG TGGGGGTCTT ATCCCTGTGC TGAGCCCTGA GGGCAGAGAG GATGGCATGT 1750
TTCAGGGGAG GGGGAAGCCT TCCTCTCAAT TTGTTGTCAG TGAAATTCCA ATAAATGGGA 1810
TTTGCTCTCT GCCT 1824

```

Sequence No.: 53

Sequence length: 1076

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP01098

Sequence characteristics

Code representing characteristics: CDS

Existence site: 62.. 601

Characterization method: E

Sequence description

```

AGTTCCGCCC GCTGGTCATC GCGCCCTTTC CCCTGCCGGT GTCCTGCTCG CCGTCCCCGC 60
C ATG CTG TCT CTA GAC TTT TTG GAC GAT GTG CGG CGG ATG AAC AAG CGG 109
Met Leu Ser Leu Asp Phe Leu Asp Asp Val Arg Arg Met Asn Lys Arg
1 5 10 15
CAG CTC TAT TAT CAA GTC CTA AAT TTT GGA ATG ATT GTC TCA TCG GCA 157
Gln Leu Tyr Tyr Gln Val Leu Asn Phe Gly Met Ile Val Ser Ser Ala
20 25 30
CTA ATG ATC TGG AAG GGG TTA ATG GTA ATA ACT GGA AGT GAA AGT CCG 205
Leu Met Ile Trp Lys Gly Leu Met Val Ile Thr Gly Ser Glu Ser Pro
35 40 45
ATT GTA GTG GTG CTC AGT GGC AGC ATG GAA CCT GCA TTT CAT AGA GGA 253
Ile Val Val Val Leu Ser Gly Ser Met Glu Pro Ala Phe His Arg Gly
50 55 60
GAT CTT CTC TTT CTA ACA AAT CGA GTT GAA GAT CCC ATA CGA GTG GGA 301
Asp Leu Leu Phe Leu Thr Asn Arg Val Glu Asp Pro Ile Arg Val Gly
65 70 75 80
GAA ATT GTT GTT TTT AGG ATA GAA GGA AGA GAG ATT CCT ATA GTT CAC 349

```

138

Glu Ile Val Val Phe Arg Ile Glu Gly Arg Glu Ile Pro Ile Val His	
85 90 95	
CGA GTC TTG AAG ATT CAT GAA AAG CAA AAT GGG CAT ATC AAG TTT TTG	397
Arg Val Leu Lys Ile His Glu Lys Gln Asn Gly His Ile Lys Phe Leu	
100 105 110	
ACC AAA GGA GAT AAT AAT GCG GTT GAT GAC CGA GGC CTC TAT AAA CAA	445
Thr Lys Gly Asp Asn Asn Ala Val Asp Asp Arg Gly Leu Tyr Lys Gln	
115 120 125	
GGA CAA CAT TGG CTA GAG AAA AAA GAT GTT GTG GGG AGA GCC AGG GGA	493
Gly Gln His Trp Leu Glu Lys Lys Asp Val Val Gly Arg Ala Arg Gly	
130 135 140	
TTT GTT CCT TAT ATT GGA ATT GTG ACG ATC CTC ATG AAT GAC TAT CCT	541
Phe Val Pro Tyr Ile Gly Ile Val Thr Ile Leu Met Asn Asp Tyr Pro	
145 150 155 160	
AAA TTT AAG TAT GCA GTT CTC TTT TTG CTG GGT TTA TTC GTG CTG GTT	589
Lys Phe Lys Tyr Ala Val Leu Phe Leu Leu Gly Leu Phe Val Leu Val	
165 170 175	
CAT CGT GAG TA AGAAGCC TGCCTTGCTG TTCCTGGGAA GAT	630
His Arg Glu	
GCCATAGTTT TCGTTACTGG ATGTTTGGAG TAGATACTGG TCTGTGATTG GTGGAATGGA	690
GAACACACGT GTTGGTGCTT CTGGGTAGCA CTGGTTTGCA TTAGTTTATG TTTCCATGCC	750
AGAGTTTGTG TGGGCGGGCG CATGTGCACC ACAGAGTGCA CTCGAGGGGA CTTTCAGTCA	810
CAGGATTTCA TAATTGTCAT TGTCACACTT TCAAATTTTT GTACATCAGT GAATTTTTTT	870
ATATTAAAAG GTTGAGCCAA AGCCCCCAGT GTTTGTATTT TGAAGCCAAG CTTCACTTCT	930
AAAGTGCTTA CAGAGACTTG TAAATGAAAA TGCAGCTCTG CACGAGTTTG AAACCGTCAT	990
ACCTCCTTCT ATTAGGAATG GCATATACTG AGGTGGTCGT AAGTCTTAAC TTCTAAAATT	1050
TTAAATAAAA GACTTTGCAC ATTGAG	1076

Sequence No.: 54

Sequence length: 1591

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Liver

Clone name: HP01148

Sequence characteristics

Code representing characteristics: CDS

Existence site: 102.. 1145

Characterization method: E



## Sequence description

```

GTCCCTCCTC TTAACATACT TGCAGCTAAA ACTAAATATT GCTGCTTGGG GACCTCCTTC      60
TAGCCTTAAA TTTCAGCTCA TCACCTTCAC CTGCCTTGGT C ATG GCT CTG CTA TTC      116
                                     Met Ala Leu Leu Phe
                                     1                               5

TCC TTG ATC CTT GCC ATT TGC ACC AGA CCT GGA TTC CTA GCG TCT CCA      164
Ser Leu Ile Leu Ala Ile Cys Thr Arg Pro Gly Phe Leu Ala Ser Pro
                                     10                               15                               20

TCT GGA GTG CGG CTG GTG GGG GGC CTC CAC CGC TGT GAA GGG CGG GTG      212
Ser Gly Val Arg Leu Val Gly Gly Leu His Arg Cys Glu Gly Arg Val
                                     25                               30                               35

GAG GTG GAA CAG AAA GGC CAG TGG GGC ACC GTG TGT GAT GAC GGC TGG      260
Glu Val Glu Gln Lys Gly Gln Trp Gly Thr Val Cys Asp Asp Gly Trp
                                     40                               45                               50

GAC ATT AAG GAC GTG GCT GTG TTG TGC CGG GAG CTG GGC TGT GGA GCT      308
Asp Ile Lys Asp Val Ala Val Leu Cys Arg Glu Leu Gly Cys Gly Ala
                                     55                               60                               65

GCC AGC GGA ACC CCT AGT GGT ATT TTG TAT GAG CCA CCA GCA GAA AAA      356
Ala Ser Gly Thr Pro Ser Gly Ile Leu Tyr Glu Pro Pro Ala Glu Lys
                                     70                               75                               80                               85

GAG CAA AAG GTC CTC ATC CAA TCA GTC AGT TGC ACA GGA ACA GAA GAT      404
Glu Gln Lys Val Leu Ile Gln Ser Val Ser Cys Thr Gly Thr Glu Asp
                                     90                               95                               100

ACA TTG GCT CAG TGT GAG CAA GAA GAA GTT TAT GAT TGT TCA CAT GAA      452
Thr Leu Ala Gln Cys Glu Gln Glu Glu Val Tyr Asp Cys Ser His Glu
                                     105                               110                               115

GAA GAT GCT GGG GCA TCG TGT GAG AAC CCA GAG AGC TCT TTC TCC CCA      500
Glu Asp Ala Gly Ala Ser Cys Glu Asn Pro Glu Ser Ser Phe Ser Pro
                                     120                               125                               130

GTC CCA GAG GGT GTC AGG CTG GCT GAC GGC CCT GGG CAT TGC AAG GGA      548
Val Pro Glu Gly Val Arg Leu Ala Asp Gly Pro Gly His Cys Lys Gly
                                     135                               140                               145

CGC GTG GAA GTG AAG CAC CAG AAC CAG TGG TAT ACC GTG TGC CAG ACA      596
Arg Val Glu Val Lys His Gln Asn Gln Trp Tyr Thr Val Cys Gln Thr
                                     150                               155                               160                               165

GGC TGG AGC CTC CGG GCC GCA AAG GTG GTG TGC CGG CAG CTG GGA TGT      644
Gly Trp Ser Leu Arg Ala Ala Lys Val Val Cys Arg Gln Leu Gly Cys
                                     170                               175                               180

GGG AGG GCT GTA CTG ACT CAA AAA CGC TGC AAC AAG CAT GCC TAT GGC      692
Gly Arg Ala Val Leu Thr Gln Lys Arg Cys Asn Lys His Ala Tyr Gly
                                     185                               190                               195

CGA AAA CCC ATC TGG CTG AGC CAG ATG TCA TGC TCA GGA CGA GAA GCA      740
Arg Lys Pro Ile Trp Leu Ser Gln Met Ser Cys Ser Gly Arg Glu Ala

```

140

200	205	210	
ACC CTT CAG GAT TGC CCT TCT GGG CCT TGG GGG AAG AAC ACC TGC AAC			788
Thr Leu Gln Asp Cys Pro Ser Gly Pro Trp Gly Lys Asn Thr Cys Asn			
215	220	225	
CAT GAT GAA GAC ACG TGG GTC GAA TGT GAA GAT CCC TTT GAC TTG AGA			836
His Asp Glu Asp Thr Trp Val Glu Cys Glu Asp Pro Phe Asp Leu Arg			
230	235	240	245
CTA GTA GGA GGA GAC AAC CTC TGC TCT GGG CGA CTG GAG GTG CTG CAC			884
Leu Val Gly Gly Asp Asn Leu Cys Ser Gly Arg Leu Glu Val Leu His			
250	255	260	
AAG GGC GTA TGG GGC TCT GTC TGT GAT GAC AAC TGG GGA GAA AAG GAG			932
Lys Gly Val Trp Gly Ser Val Cys Asp Asp Asn Trp Gly Glu Lys Glu			
265	270	275	
GAC CAG GTG GTA TGC AAG CAA CTG GGC TGT GGG AAG TCC CTC TCT CCC			980
Asp Gln Val Val Cys Lys Gln Leu Gly Cys Gly Lys Ser Leu Ser Pro			
280	285	290	
TCC TTC AGA GAC CGG AAA TGC TAT GGC CCT GGG GTT GGC CGC ATC TGG			1028
Ser Phe Arg Asp Arg Lys Cys Tyr Gly Pro Gly Val Gly Arg Ile Trp			
295	300	305	
CTG GAT AAT GTT CGT TGC TCA GGG GAG GAG CAG TCC CTG GAG CAG TGC			1076
Leu Asp Asn Val Arg Cys Ser Gly Glu Glu Gln Ser Leu Glu Gln Cys			
310	315	320	325
CAG CAC AGA TTT TGG GGG TTT CAC GAC TGC ACC CAC CAG GAA GAT GTG			1124
Gln His Arg Phe Trp Gly Phe His Asp Cys Thr His Gln Glu Asp Val			
330	335	340	
GCT GTC ATC TGC TCA GGA TAGTATCCTG GTGTTGCTTG ACCTGGCC			1170
Ala Val Ile Cys Ser Gly			
345			
CCCCTGGCCC CGCCTGCCCT CTGCTTGTTT TCCTGAGCCC TGATTATCCT CATACTCATT			1230
CTGGGGCTCA GGCTTGAGCC ACTACTCCCT CATCCCCTCA GGAGTCTGAA CACTGGGCTT			1290
ATGCCTTACT CTCAGGGACA AGCAGCCCCC ATTGCTGCCT GTAGATGTGA GCTGTTGAGT			1350
TCCCTCTTGC TGGGGAAGAT GAGCTTCCAT GTATCCTGTG CTCAACCCTG ACCCTTTGAC			1410
ACTGTTCTG GCCTTTCCTG CCTTTTCTCA AGCTGCCTGG AATCCTCAAA CCTGTCACTT			1470
TGGTCAGATG TGCAGACCAT TACTAAGGTC TATGTCTGCA AACATTACTA ATCTAGGTCC			1530
TATTACTAAT CTATGTCTGC AAACATTAAA GGAATGAAAC AATGAAAGGA ACATTTGAAA			1590
G			1591

Sequence No.: 55

Sequence length: 1888

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

## Original source:

Organism species: *Homo sapiens*

Cell kind: Liver

Clone name: HP01293

## Sequence characteristics

Code representing characteristics: CDS

Existence site: 90.. 1754

Characterization method: E

## Sequence description

```

CCTTTTCAAA GATCTCTGAG GGAGACATTG CACCTGGCCA CTGCAGCCCA GAGCAGGTCT      60
GGCCACGGCC ATGAGCATGC TGAGCCATC ATG CCC ACC GTG GAT GAC ATT CTG      113
                               Met Pro Thr Val Asp Asp Ile Leu
                               1           5

GAG CAG GTT GGG GAG TCT GGC TGG TTC CAG AAG CAA GCC TTC CTC ATC      161
Glu Gln Val Gly Glu Ser Gly Trp Phe Gln Lys Gln Ala Phe Leu Ile
    10           15           20

TTA TGC CTG CTG TCG GCT GCC TTT GCG CCC ATC TGT GTG GGC ATC GTC      209
Leu Cys Leu Leu Ser Ala Ala Phe Ala Pro Ile Cys Val Gly Ile Val
    25           30           35           40

TTC CTG GGT TTC ACA CCT GAC CAC CAC TGC CAG AGT CCT GGG GTG GCT      257
Phe Leu Gly Phe Thr Pro Asp His His Cys Gln Ser Pro Gly Val Ala
           45           50           55

GAG CTG AGC CAG CGC TGT GGC TGG AGC CCT GCG GAG GAG CTG AAC TAT      305
Glu Leu Ser Gln Arg Cys Gly Trp Ser Pro Ala Glu Glu Leu Asn Tyr
           60           65           70

ACA GTG CCA GGC CTG GGG CCC GCG GGC GAG GCC TTC CTT GGC CAG TGC      353
Thr Val Pro Gly Leu Gly Pro Ala Gly Glu Ala Phe Leu Gly Gln Cys
           75           80           85

AGG CGC TAT GAA GTG GAC TGG AAC CAG AGC GCC CTC AGC TGT GTA GAC      401
Arg Arg Tyr Glu Val Asp Trp Asn Gln Ser Ala Leu Ser Cys Val Asp
           90           95           100

CCC CTG GCT AGC CTG GCC ACC AAC AGG AGC CAC CTG CCG CTG GGT CCC      449
Pro Leu Ala Ser Leu Ala Thr Asn Arg Ser His Leu Pro Leu Gly Pro
    105           110           115           120

TGC CAG GAT GGC TGG GTG TAT GAC ACG CCC GGC TCT TCC ATC GTC ACT      497
Cys Gln Asp Gly Trp Val Tyr Asp Thr Pro Gly Ser Ser Ile Val Thr
           125           130           135

GAG TTC AAC CTG GTG TGT GCT GAC TCC TGG AAG CTG GAC CTC TTT CAG      545
Glu Phe Asn Leu Val Cys Ala Asp Ser Trp Lys Leu Asp Leu Phe Gln
           140           145           150

TCC TGT TTG AAT GCG GGC TTC TTC TTT GGC TCT CTC GGT GTT GGC TAC      593
Ser Cys Leu Asn Ala Gly Phe Phe Phe Gly Ser Leu Gly Val Gly Tyr
           155           160           165

```

TTT GCA GAC AGG TTT GGC CGT AAG CTG TGT CTC CTG GGA ACT GTG CTG	641
Phe Ala Asp Arg Phe Gly Arg Lys Leu Cys Leu Leu Gly Thr Val Leu	
170 175 180	
GTC AAC GCG GTG TCG GGC GTG CTC ATG GCC TTC TCG CCC AAC TAC ATG	689
Val Asn Ala Val Ser Gly Val Leu Met Ala Phe Ser Pro Asn Tyr Met	
185 190 195 200	
TCC ATG CTG CTC TTC CGC CTG CTG CAG GGC CTG GTC AGC AAG GGC AAC	737
Ser Met Leu Leu Phe Arg Leu Leu Gln Gly Leu Val Ser Lys Gly Asn	
205 210 215	
TGG ATG GCT GGC TAC ACC CTA ATC ACA GAA TTT GTT GGC TCG GGC TCC	785
Trp Met Ala Gly Tyr Thr Leu Ile Thr Glu Phe Val Gly Ser Gly Ser	
220 225 230	
AGA AGA ACG GTG GCG ATC ATG TAC CAG ATG GCC TTC ACG GTG GGG CTG	833
Arg Arg Thr Val Ala Ile Met Tyr Gln Met Ala Phe Thr Val Gly Leu	
235 240 245	
GTG GCG CTT ACC GGG CTG GCC TAC GCC CTG CCT CAC TGG CGC TGG CTG	881
Val Ala Leu Thr Gly Leu Ala Tyr Ala Leu Pro His Trp Arg Trp Leu	
250 255 260	
CAG CTG GCA GTC TCC CTG CCC ACC TTC CTC TTC CTG CTC TAC TAC TGG	929
Gln Leu Ala Val Ser Leu Pro Thr Phe Leu Phe Leu Leu Tyr Tyr Trp	
265 270 275 280	
TGT GTG CCG GAG TCC CCT CGG TGG CTG TTA TCA CAA AAA AGA AAC ACT	977
Cys Val Pro Glu Ser Pro Arg Trp Leu Leu Ser Gln Lys Arg Asn Thr	
285 290 295	
GAA GCA ATA AAG ATA ATG GAC CAC ATC GCT CAA AAG AAT GGG AAG TTG	1025
Glu Ala Ile Lys Ile Met Asp His Ile Ala Gln Lys Asn Gly Lys Leu	
300 305 310	
CCT CCT GCT GAT TTA AAG ATG CTT TCC CTC GAA GAG GAT GTC ACC GAA	1073
Pro Pro Ala Asp Leu Lys Met Leu Ser Leu Glu Glu Asp Val Thr Glu	
315 320 325	
AAG CTG AGC CCT TCA TTT GCA GAC CTG TTC CGC ACG CCG CGC CTG AGG	1121
Lys Leu Ser Pro Ser Phe Ala Asp Leu Phe Arg Thr Pro Arg Leu Arg	
330 335 340	
AAG CGC ACC TTC ATC CTG ATG TAC CTG TGG TTC ACG GAC TCT GTG CTC	1169
Lys Arg Thr Phe Ile Leu Met Tyr Leu Trp Phe Thr Asp Ser Val Leu	
345 350 355 360	
TAT CAG GGG CTC ATC CTG CAC ATG GGC GCC ACC AGC GGG AAC CTC TAC	1217
Tyr Gln Gly Leu Ile Leu His Met Gly Ala Thr Ser Gly Asn Leu Tyr	
365 370 375	
CTG GAT TTC CTT TAC TCC GCT CTG GTC GAA ATC CCG GGG GCC TTC ATA	1265
Leu Asp Phe Leu Tyr Ser Ala Leu Val Glu Ile Pro Gly Ala Phe Ile	
380 385 390	
GCC CTC ATC ACC ATT GAC CGC GTG GGC CGC ATC TAC CCC ATG GCC GTG	1313
Ala Leu Ile Thr Ile Asp Arg Val Gly Arg Ile Tyr Pro Met Ala Val	

395	400	405	
TCA AAT TTG TTG GCG GGG GCA GCC TGC CTC GTC ATG ATT TTT ATC TCA			1361
Ser Asn Leu Leu Ala Gly Ala Ala Cys Leu Val Met Ile Phe Ile Ser			
410	415	420	
CCT GAC CTG CAC TGG TTA AAC ATC ATA ATC ATG TGT GTT GGC CGA ATG			1409
Pro Asp Leu His Trp Leu Asn Ile Ile Ile Met Cys Val Gly Arg Met			
425	430	435	440
GGA ATC ACC ATT GCA ATA CAA ATG ATC TGC CTG GTG AAT GCT GAG CTG			1457
Gly Ile Thr Ile Ala Ile Gln Met Ile Cys Leu Val Asn Ala Glu Leu			
445	450	455	
TAC CCC ACA TTC GTC AGG AAC CTC GGA GTG ATG GTG TGT TCC TCC CTG			1505
Tyr Pro Thr Phe Val Arg Asn Leu Gly Val Met Val Cys Ser Ser Leu			
460	465	470	
TGT GAC ATA GGT GGG ATA ATC ACC CCC TTC ATA GTC TTC AGG CTG AGG			1553
Cys Asp Ile Gly Gly Ile Ile Thr Pro Phe Ile Val Phe Arg Leu Arg			
475	480	485	
GAG GTC TGG CAA GCC TTG CCC CTC ATT TTG TTT GCG GTG TTG GGC CTG			1601
Glu Val Trp Gln Ala Leu Pro Leu Ile Leu Phe Ala Val Leu Gly Leu			
490	495	500	
CTT GCC GCG GGA GTG ACG CTA CTT CTT CCA GAG ACC AAG GGG GTC GCT			1649
Leu Ala Ala Gly Val Thr Leu Leu Leu Pro Glu Thr Lys Gly Val Ala			
505	510	515	520
TTG CCA GAG ACC ATG AAG GAC GCC GAG AAC CTT GGG AGA AAA GCA AAG			1697
Leu Pro Glu Thr Met Lys Asp Ala Glu Asn Leu Gly Arg Lys Ala Lys			
525	530	535	
CCC AAA GAA AAC ACG ATT TAC CTT AAG GTC CAA ACC TCA GAA CCC TCG			1745
Pro Lys Glu Asn Thr Ile Tyr Leu Lys Val Gln Thr Ser Glu Pro Ser			
540	545	550	
GGC ACC TGAGAGAGAT GTTTTGCGGC GATGTCGTGT TGGAGGGATG AAGATGGAG			1800
Gly Thr			
TTATCCTCTG CAGAAATTCC TAGACGCCTT CACTTCTCTG TATTCTTCCT CATACTTGCC			1860
TACCCCAAAA TTAATATCAG TCCTAAAG			1888

Sequence No.: 56

Sequence length: 2033

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP10013

## Sequence characteristics

Code representing characteristics: CDS

Existence site: 97.. 1149

Characterization method: E

## Sequence description

GAGTCCGAGC GCGTCACCTC CTCACGCTGC GGCTGTCGCC CGTGTCCCGC CGGCCCGTTC	60
CGTGTGCGCC CGCAGTGCTG CGGCCGCGCC GGCACC ATG GCT GTG TTT GTC GTG	114
Met Ala Val Phe Val Val	
1 5	
CTC CTG GCG TTG GTG GCG GGT GTT TTG GGG AAC GAG TTT AGT ATA TTA	162
Leu Leu Ala Leu Val Ala Gly Val Leu Gly Asn Glu Phe Ser Ile Leu	
10 15 20	
AAA TCA CCA GGG TCT GTT GTT TTC CGA AAT GGA AAT TGG CCT ATA CCA	210
Lys Ser Pro Gly Ser Val Val Phe Arg Asn Gly Asn Trp Pro Ile Pro	
25 30 35	
GGA GAG CGG ATC CCA GAC GTG GCT GCA TTG TCC ATG GGC TTC TCT GTG	258
Gly Glu Arg Ile Pro Asp Val Ala Ala Leu Ser Met Gly Phe Ser Val	
40 45 50	
AAA GAA GAC CTT TCT TGG CCA GGA CTC GCA GTG GGT AAC CTG TTT CAT	306
Lys Glu Asp Leu Ser Trp Pro Gly Leu Ala Val Gly Asn Leu Phe His	
55 60 65 70	
CGT CCT CGG GCT ACC GTC ATG GTG ATG GTG AAG GGA GTG AAC AAA CTG	354
Arg Pro Arg Ala Thr Val Met Val Met Val Lys Gly Val Asn Lys Leu	
75 80 85	
GCT CTA CCC CCA GGC AGT GTC ATT TCG TAC CCT TTG GAG AAT GCA GTT	402
Ala Leu Pro Pro Gly Ser Val Ile Ser Tyr Pro Leu Glu Asn Ala Val	
90 95 100	
CCT TTT AGT CTT GAC AGT GTT GCA AAT TCC ATT CAC TCC TTA TTT TCT	450
Pro Phe Ser Leu Asp Ser Val Ala Asn Ser Ile His Ser Leu Phe Ser	
105 110 115	
GAG GAA ACT CCT GTT GTT TTG CAG TTG GCT CCC AGT GAG GAA AGA GTG	498
Glu Glu Thr Pro Val Val Leu Gln Leu Ala Pro Ser Glu Glu Arg Val	
120 125 130	
TAT ATG GTA GGG AAG GCA AAC TCA GTG TTT GAA GAC CTT TCA GTC ACC	546
Tyr Met Val Gly Lys Ala Asn Ser Val Phe Glu Asp Leu Ser Val Thr	
135 140 145 150	
TTG CGC CAG CTC CGT AAT CGC CTG TTT CAA GAA AAC TCT GTT CTC AGT	594
Leu Arg Gln Leu Arg Asn Arg Leu Phe Gln Glu Asn Ser Val Leu Ser	
155 160 165	
TCA CTC CCC CTC AAT TCT CTG AGT AGG AAC AAT GAA GTT GAC CTG CTC	642
Ser Leu Pro Leu Asn Ser Leu Ser Arg Asn Asn Glu Val Asp Leu Leu	

170	175	180	
TTT CTT TCT GAA CTG CAA GTG CTA CAT GAT ATT TCA AGC TTG CTG TCT			690
Phe Leu Ser Glu Leu Gln Val Leu His Asp Ile Ser Ser Leu Leu Ser			
185	190	195	
CGT CAT AAG CAT CTA GCC AAG GAT CAT TCT CCT GAT TTA TAT TCA CTG			738
Arg His Lys His Leu Ala Lys Asp His Ser Pro Asp Leu Tyr Ser Leu			
200	205	210	
GAG CTG GCA GGT TTG GAT GAA ATT GGG AAG CGT TAT GGG GAA GAC TCT			786
Glu Leu Ala Gly Leu Asp Glu Ile Gly Lys Arg Tyr Gly Glu Asp Ser			
215	220	225	230
GAA CAA TTC AGA GAT GCT TCT AAG ATC CTT GTT GAC GCT CTG CAA AAG			834
Glu Gln Phe Arg Asp Ala Ser Lys Ile Leu Val Asp Ala Leu Gln Lys			
235	240	245	
TTT GCA GAT GAC ATG TAC AGT CTT TAT GGT GGG AAT GCA GTG GTA GAG			882
Phe Ala Asp Asp Met Tyr Ser Leu Tyr Gly Gly Asn Ala Val Val Glu			
250	255	260	
TTA GTC ACT GTC AAG TCA TTT GAC ACC TCC CTC ATT AGG AAG ACA AGG			930
Leu Val Thr Val Lys Ser Phe Asp Thr Ser Leu Ile Arg Lys Thr Arg			
265	270	275	
ACT ATC CTT GAG GCA AAA CAA GCG AAG AAC CCA GCA AGT CCC TAT AAC			978
Thr Ile Leu Glu Ala Lys Gln Ala Lys Asn Pro Ala Ser Pro Tyr Asn			
280	285	290	
CTT GCA TAT AAG TAT AAT TTT GAA TAT TCC GTG GTT TTC AAC ATG GTA			1026
Leu Ala Tyr Lys Tyr Asn Phe Glu Tyr Ser Val Val Phe Asn Met Val			
295	300	305	310
CTT TGG ATA ATG ATC GCC TTG GCC TTG GCT GTG ATT ATC ACC TCT TAC			1074
Leu Trp Ile Met Ile Ala Leu Ala Leu Ala Val Ile Ile Thr Ser Tyr			
315	320	325	
AAT ATT TGG AAC ATG GAT CCT GGA TAT GAT AGC ATC ATT TAT AGG ATG			1122
Asn Ile Trp Asn Met Asp Pro Gly Tyr Asp Ser Ile Ile Tyr Arg Met			
330	335	340	
ACA AAC CAG AAG ATT CGA ATG GAT TGAATGTTAC CTGTGCCAGA ATTA			1170
Thr Asn Gln Lys Ile Arg Met Asp			
345	350		
GAAAAGGGGG TTGGAAATTG GCTGTTTTGT TAAAATATAT CTTTGTAGTGT GCTTTAAAGT			1230
AGATAGTATA CTTTACATTT ATAAAAAAA ATCAAATTTT GTTCTTTATT TTGTGTGTGC			1290
CTGTGATGTT TTTCTAGAGT GAATTATAGT ATTGACGTGA ATCCCACTGT GGTATAGATT			1350
CCATAATATG CTTGAATATT ATGATATAGC CATTTAATAA CATTGATTTC ATTCTGTTTA			1410
ATGAATTTGG AAATATGCAC TGAAAGAAAT GTAAAACATT TAGAATAGCT CGTGTATATGG			1470
AAAAAAGTGC ACTGAATTTA TTAGACAAAC TTACGAATGC TTAACCTCTT TACACAGCAT			1530
AGCTGAAAAT CATATTTGGG CTATTGTATA CTATGAACAA TTTGTAAATG TCTTAATTTG			1590
ATGTAAATAA CTCTGAAACA AGAGAAAAGG TTTTAACTT AGAGTAGCCC TAAAATATGG			1650
ATGTGCTTAT ATAATCGCTT AGTTTTGGAA CTGTATCTGA GTAACAGAGG ACAGCTGTTT			1710
TTTAAACCTC TTCTGCAAGT TTGTTGACCT ACATGGGCTA ATATGGATAC TAAAAATACT			1770

ACATTGATCT AAGAAGAAAC TAGCCTTG TG GAGTATATAG ATGCTTTTCA TTATACACAC 1830  
 AAAAATCCCT GAGGGACATT TTGAGGCATG AATATAAAAC ATTTTATTT CAGTAACTTT 1890  
 TCCCCCTGTG TAAGTTACTA TGGTTTGTGG TACAACTTCA TTCTATACAA TATTAAGTGG 1950  
 AAGTGGGTGA ATTCTACTTT TTATGTTGGA GTGGACCAAT GTCTATCAAG AGTGACAAAT 2010  
 AAAGTTAATG ATGATTCCAA AAC 2033

Sequence No.: 57

Sequence length: 911

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Fibrosarcoma

Cell line: HT-1080

Clone name: HP10034

Sequence characteristics

Code representing characteristics: CDS

Existence site: 176.. 805

Characterization method: E

Sequence description

ACGCCTGGGT GACCTCTACG TATATACAGA GCCTCCCTGG CCCTCCTGGA AAGAGTCCTG 60  
 GAAAGACAAC CTTCAGGTCC AGCCCTGGAG CTGGAGGAGT GGAGCCCCAC TCTGAAGACG 120  
 CAGCCTTTCT CCAGGTTCTG TCTCTCCCAT TCTGATTCTT GACACCAGAT GCAGG ATG 178  
 Met  
 1  
 GTG TCC TCT CCC TGC ACG CAG GCA AGC TCA CGG ACT TGC TCC CGT ATC 226  
 Val Ser Ser Pro Cys Thr Gln Ala Ser Ser Arg Thr Cys Ser Arg Ile  
 5 10 15  
 CTG GGA CTG AGC CTT GGG ACT GCA GCC CTG TTT GCT GCT GGG GCC AAC 274  
 Leu Gly Leu Ser Leu Gly Thr Ala Ala Leu Phe Ala Ala Gly Ala Asn  
 20 25 30  
 GTG GCA CTC CTC CTT CCT AAC TGG GAT GTC ACC TAC CTG TTG AGG GGC 322  
 Val Ala Leu Leu Leu Pro Asn Trp Asp Val Thr Tyr Leu Leu Arg Gly  
 35 40 45  
 CTC CTT GGC AGG CAT GCC ATG CTG GGA ACT GGG CTC TGG GGA GGA GGC 370  
 Leu Leu Gly Arg His Ala Met Leu Gly Thr Gly Leu Trp Gly Gly Gly  
 50 55 60 65  
 CTC ATG GTA CTC ACT GCA GCT ATC CTC ATC TCC TTG ATG GGC TGG AGA 418  
 Leu Met Val Leu Thr Ala Ala Ile Leu Ile Ser Leu Met Gly Trp Arg



70	75	80	
TAC GGC TGC TTC AGT AAG AGT GGG CTC TGT CGA AGC GTG CTT ACT GCT			466
Tyr Gly Cys Phe Ser Lys Ser Gly Leu Cys Arg Ser Val Leu Thr Ala			
85	90	95	
CTG TTG TCA GGT GGC CTG GCT TTA CTT GGA GCC CTG ATT TGC TTT GTC			514
Leu Leu Ser Gly Gly Leu Ala Leu Leu Gly Ala Leu Ile Cys Phe Val			
100	105	110	
ACT TCT GGA GTT GCT CTG AAA GAT GGT CCT TTT TGC ATG TTT GAT GTT			562
Thr Ser Gly Val Ala Leu Lys Asp Gly Pro Phe Cys Met Phe Asp Val			
115	120	125	
TCA TCC TTC AAT CAG ACA CAA GCT TGG AAA TAT GGT TAC CCA TTC AAA			610
Ser Ser Phe Asn Gln Thr Gln Ala Trp Lys Tyr Gly Tyr Pro Phe Lys			
130	135	140	145
GAC CTG CAT AGT AGG AAT TAT CTG TAT GAC CGT TCG CTC TGG AAC TCC			658
Asp Leu His Ser Arg Asn Tyr Leu Tyr Asp Arg Ser Leu Trp Asn Ser			
150	155	160	
GTC TGC CTG GAG CCC TCT GCA GCT GTT GTC TGG CAC GTG TCC CTC TTC			706
Val Cys Leu Glu Pro Ser Ala Ala Val Val Trp His Val Ser Leu Phe			
165	170	175	
TCC GCC CTT CTG TGC ATC AGC CTG CTC CAG CTT CTC CTG GTG GTC GTT			754
Ser Ala Leu Leu Cys Ile Ser Leu Leu Gln Leu Leu Leu Val Val Val			
180	185	190	
CAT GTC ATC AAC AGC CTC CTG GGC CTT TTC TGC AGC CTC TGC GAG AAG			802
His Val Ile Asn Ser Leu Leu Gly Leu Phe Cys Ser Leu Cys Glu Lys			
195	200	205	
TGACAGGC AGAACCTTCA CTTGCAAGCA TGGGTGTTA TCATCATCGG CTGTCTTGAA			860
TCCTTTCTAC AAGGAGTGGG TACGAATTAT AAACAAACTT CCCCTTTAGG T			911

Sequence No.: 58

Sequence length: 601

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Fibrosarcoma

Cell line: HT-1080

Clone name: HP10050

Sequence characteristics

Code representing characteristics: CDS

Existence site: 10.. 501

Characterization method: E

## Sequence description

```

CCATCTGTC ATG GCG GCT GGG CTG TTT GGT TTG AGC GCT CGC CGT CTT TTG      51
      Met Ala Ala Gly Leu Phe Gly Leu Ser Ala Arg Arg Leu Leu
          1           5           10
GCG GCA GCG GCG ACG CGA GGG CTC CCG GCC GCC CGC GTC CGC TGG GAA      99
Ala Ala Ala Ala Thr Arg Gly Leu Pro Ala Ala Arg Val Arg Trp Glu
      15           20           25           30
TCT AGC TTC TCC AGG ACT GTG GTC GCC CCG TCC GCT GTG GCG GGA AAG      147
Ser Ser Phe Ser Arg Thr Val Val Ala Pro Ser Ala Val Ala Gly Lys
          35           40           45
CGG CCC CCA GAA CCG ACC ACA CCG TGG CAA GAG GAC CCA GAA CCC GAG      195
Arg Pro Pro Glu Pro Thr Thr Pro Trp Gln Glu Asp Pro Glu Pro Glu
          50           55           60
GAC GAA AAC TTG TAT GAG AAG AAC CCA GAC TCC CAT GGT TAT GAC AAG      243
Asp Glu Asn Leu Tyr Glu Lys Asn Pro Asp Ser His Gly Tyr Asp Lys
          65           70           75
GAC CCC GTT TTG GAC GTC TGG AAC ATG CGA CTT GTC TTC TTC TTT GGC      291
Asp Pro Val Leu Asp Val Trp Asn Met Arg Leu Val Phe Phe Phe Gly
          80           85           90
GTC TCC ATC ATC CTG GTC CTT GGC AGC ACC TTT GTG GCC TAT CTG CCT      339
Val Ser Ile Ile Leu Val Leu Gly Ser Thr Phe Val Ala Tyr Leu Pro
          95          100          105          110
GAC TAC AGG TGC ACA GGG TGT CCA AGA GCG TGG GAT GGG ATG AAA GAG      387
Asp Tyr Arg Cys Thr Gly Cys Pro Arg Ala Trp Asp Gly Met Lys Glu
          115          120          125
TGG TCC CGC CGC GAA GCT GAG AGG CTT GTG AAA TAC CGA GAG GCC AAT      435
Trp Ser Arg Arg Glu Ala Glu Arg Leu Val Lys Tyr Arg Glu Ala Asn
          130          135          140
GGC CTT CCC ATC ATG GAA TCC AAC TGC TTC GAC CCC AGC AAG ATC CAG      483
Gly Leu Pro Ile Met Glu Ser Asn Cys Phe Asp Pro Ser Lys Ile Gln
          145          150          155
CTG CCA GAG GAT GAG TGACCAGTTG CTAAGTGGGG CTC AAGAAGC AC      530
Leu Pro Glu Asp Glu
          160
CGCCTTCCCC ACCCCCTGCC TGCCATTCTG ACCTCTTCTC AGAGCACCTA ATTAAAGGGG      590
CTGAAAGTCT G      601

```

Sequence No.: 59

Sequence length: 394

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10071

Sequence characteristics

Code representing characteristics: CDS

Existence site: 47.. 325

Characterization method: E

Sequence description

```

AACATCCGGG CCGCGCGGGG AAGGGGAGAC GTGGGTAGA GTGACC ATG ACG AAA      55
                                     Met Thr Lys
                                     1
TTA GCG CAG TGG CTT TGG GGA CTA GCG ATC CTG GGC TCC ACC TGG GTG      103
Leu Ala Gln Trp Leu Trp Gly Leu Ala Ile Leu Gly Ser Thr Trp Val
      5              10              15
GCC CTG ACC ACG GGA GCC TTG GGC CTG GAG CTG CCC TTG TCC TGC CAG      151
Ala Leu Thr Thr Gly Ala Leu Gly Leu Glu Leu Pro Leu Ser Cys Gln
      20              25              30              35
GAA GTC CTG TGG CCA CTG CCC GCC TAC TTG CTG GTG TCC GCC GGC TGC      199
Glu Val Leu Trp Pro Leu Pro Ala Tyr Leu Leu Val Ser Ala Gly Cys
              40              45              50
TAT GCC CTG GGC ACT GTG GGC TAT CGT GTG GCC ACT TTT CAT GAC TGC      247
Tyr Ala Leu Gly Thr Val Gly Tyr Arg Val Ala Thr Phe His Asp Cys
              55              60              65
GAG GAC GCC GCA CGC GAG CTG CAG AGC CAG ATA CAG GAG GCC CGA GCC      295
Glu Asp Ala Ala Arg Glu Leu Gln Ser Gln Ile Gln Glu Ala Arg Ala
              70              75              80
GAC TTA GCC CGC AGG GGG CTG CGC TTC TGACAGCCTA ACCCCATT      340
Asp Leu Ala Arg Arg Gly Leu Arg Phe
      85              90
CCTGTGCGGA CAGCCCTTCC TCCCATTTCC CATTAAAGAG CCAGTTTATT TTCT      394

```

Sequence No.: 60

Sequence length: 732

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Lymphoma

Cell line: U937

Clone name: HP10076

## Sequence characteristics

Code representing characteristics: CDS

Existence site: 82.. 600

Characterization method: E

## Sequence description

```

AGAAACGTGT TCGCTGCCCA GAAGAAGGGA AGGCGCGAGT GAGGAAAGGA GGTACTGTAG      60
ATGCCCTCCA AATCCTTGGT T ATG GAA TAT TTG GCT CAT CCC AGT ACA CTC      111
      Met Glu Tyr Leu Ala His Pro Ser Thr Leu
              1              5              10
GGC TTG GCT GTT GGA GTT GCT TGT GGC ATG TGC CTG GGC TGG AGC CTT      159
Gly Leu Ala Val Gly Val Ala Cys Gly Met Cys Leu Gly Trp Ser Leu
              15              20              25
CGA GTA TGC TTT GGG ATG CTC CCC AAA AGC AAG ACG AGC AAG ACA CAC      207
Arg Val Cys Phe Gly Met Leu Pro Lys Ser Lys Thr Ser Lys Thr His
              30              35              40
ACA GAT ACT GAA AGT GAA GCA AGC ATC TTG GGA GAC AGC GGG GAG TAC      255
Thr Asp Thr Glu Ser Glu Ala Ser Ile Leu Gly Asp Ser Gly Glu Tyr
              45              50              55
AAG ATG ATT CTT GTG GTT CGA AAT GAC TTA AAG ATG GGA AAA GGG AAA      303
Lys Met Ile Leu Val Val Arg Asn Asp Leu Lys Met Gly Lys Gly Lys
              60              65              70
GTG GCT GCC CAG TGC TCT CAT GCT GCT GTT TCA GCC TAC AAG CAG ATT      351
Val Ala Ala Gln Cys Ser His Ala Ala Val Ser Ala Tyr Lys Gln Ile
              75              80              85              90
CAA AGA AGA AAT CCT GAA ATG CTC AAA CAA TGG GAA TAC TGT GGC CAG      399
Gln Arg Arg Asn Pro Glu Met Leu Lys Gln Trp Glu Tyr Cys Gly Gln
              95              100              105
CCC AAG GTG GTG GTC AAA GCT CCT GAT GAA GAA ACC CTG ATT GCA TTA      447
Pro Lys Val Val Val Lys Ala Pro Asp Glu Glu Thr Leu Ile Ala Leu
              110              115              120
TTG GCC CAT GCA AAA ATG CTG GGA CTG ACT GTA AGT TTA ATT CAA GAT      495
Leu Ala His Ala Lys Met Leu Gly Leu Thr Val Ser Leu Ile Gln Asp
              125              130              135
GCT GGA CGT ACT CAG ATT GCA CCA GGC TCT CAA ACT GTC CTA GGG ATT      543
Ala Gly Arg Thr Gln Ile Ala Pro Gly Ser Gln Thr Val Leu Gly Ile
              140              145              150
GGG CCA GGA CCA GCA GAC CTA ATT GAC AAA GTC ACT GGT CAC CTA AAA      591
Gly Pro Gly Pro Ala Asp Leu Ile Asp Lys Val Thr Gly His Leu Lys
              155              160              165              170
CTT TAC TAGGTGGACT TTGATATGAC AACAACCCCT CCATCACAAG TGT      640
Leu Tyr

```

TTGAAGCCTG TCAGATTCTA ACAACAAAAG CTGAATTTCT TCACCCAACT TAAATGTTCT 700  
 TGAGATGAAA ATAAACCTA TTCCCATGTT CT 732

Sequence No.: 61

Sequence length: 697

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Lymphoma

Cell line: U937

Clone name: HP10085

Sequence characteristics

Code representing characteristics: CDS

Existence site: 151.. 600

Characterization method: E

Sequence description

TATACCTCTA GTTTGGAGCT GTGCTGTAAA AACAAGAGTA ACATTTTAT ATTAAAGTTA 60  
 AATAAGTTA CAACTTTGAA GAGAGTTTCT GCAAGACATG ACACAAAGCT GCTAGCAGAA 120  
 AATCAAAACG CTGATTAAAA GAAGCACGGT ATG ATG ACC AAA CAT AAA AAG TGT 174  
 Met Met Thr Lys His Lys Lys Cys  
 1 5  
 TTT ATA ATT GTT GGT GTT TTA ATA ACA ACT AAT ATT ATT ACT CTG ATA 222  
 Phe Ile Ile Val Gly Val Leu Ile Thr Thr Asn Ile Ile Thr Leu Ile  
 10 15 20  
 GTT AAA CTA ACT CGA GAT TCT CAG AGT TTA TGC CCC TAT GAT TGG ATT 270  
 Val Lys Leu Thr Arg Asp Ser Gln Ser Leu Cys Pro Tyr Asp Trp Ile  
 25 30 35 40  
 GGT TTC CAA AAC AAA TGC TAT TAT TTC TCT AAA GAA GAA GGA GAT TGG 318  
 Gly Phe Gln Asn Lys Cys Tyr Tyr Phe Ser Lys Glu Glu Gly Asp Trp  
 45 50 55  
 AAT TCA AGT AAA TAC AAC TGT TCC ACT CAA CAT GCC GAC CTA ACT ATA 366  
 Asn Ser Ser Lys Tyr Asn Cys Ser Thr Gln His Ala Asp Leu Thr Ile  
 60 65 70  
 ATT GAC AAC ATA GAA GAA ATG AAT TTT CTT AGG CGG TAT AAA TGC AGT 414  
 Ile Asp Asn Ile Glu Glu Met Asn Phe Leu Arg Arg Tyr Lys Cys Ser  
 75 80 85  
 TCT GAT CAC TGG ATT GGA CTG AAG ATG GCA AAA AAT CGA ACA GGA CAA 462  
 Ser Asp His Trp Ile Gly Leu Lys Met Ala Lys Asn Arg Thr Gly Gln

152

90	95	100	
TGG GTA GAT GGA GCT ACA TTT ACC AAA TCG TTT GGC ATG AGA GGG AGT			510
Trp Val Asp Gly Ala Thr Phe Thr Lys Ser Phe Gly Met Arg Gly Ser			
105	110	115	120
GAA GGA TGT GCC TAC CTC AGC GAT GAT GGT GCA GCA ACA GCT AGA TGT			558
Glu Gly Cys Ala Tyr Leu Ser Asp Asp Gly Ala Ala Thr Ala Arg Cys			
	125	130	135
TAC ACC GAA AGA AAA TGG ATT TGC AGG AAA AGA ATA CAC TAA			600
Tyr Thr Glu Arg Lys Trp Ile Cys Arg Lys Arg Ile His			
	140	145	
GTTAATGTCT AAGATAATGG GGAAAATAGA AAATAACATT ATTAAGTGTA AAACCAGCAA			660
AGTACTTTTT TAATTAAACA AAGTTCGAGT TTTGTAC			697

Sequence No.: 62

Sequence length: 1186

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10122

Sequence characteristics

Code representing characteristics: CDS

Existence site: 139.. 705

Characterization method: E

Sequence description

AAGTGGATC TTCGGGCTGT CAGAGTTGGT CTGTTACTCG GTGGTGGCGG AGTCTACGGA	60
AGCCGTTTTTC GCTTCACTTT TCCTGGCTGT AGAGCGCTTT CCCCCTGGCG GGTGAGAGTG	120
CAGAGACGAA GGTGCGAG ATG AGC ACT ATG TTC GCG GAC ACT CTC CTC ATC	171
Met Ser Thr Met Phe Ala Asp Thr Leu Leu Ile	
1 5 10	
GTT TTT ATC TCT GTG TGC ACG GCT CTG CTC GCA GAG GGC ATA ACC TGG	219
Val Phe Ile Ser Val Cys Thr Ala Leu Leu Ala Glu Gly Ile Thr Trp	
15 20 25	
GTC CTG GTT TAC AGG ACA GAC AAG TAC AAG AGA CTG AAG GCA GAA GTG	267
Val Leu Val Tyr Arg Thr Asp Lys Tyr Lys Arg Leu Lys Ala Glu Val	
30 35 40	
GAA AAA CAG AGT AAA AAA TTG GAA AAG AAG AAG GAA ACA ATA ACA GAG	315
Glu Lys Gln Ser Lys Lys Leu Glu Lys Lys Lys Glu Thr Ile Thr Glu	
45 50 55	

TCA GCT GGT CGA CAA CAG AAA AAG AAA ATA GAG AGA CAA GAA GAG AAA	363
Ser Ala Gly Arg Gln Gln Lys Lys Lys Ile Glu Arg Gln Glu Glu Lys	
60 65 70 75	
CTG AAG AAT AAC AAC AGA GAT CTA TCA ATG GTT CGA ATG AAA TCC ATG	411
Leu Lys Asn Asn Asn Arg Asp Leu Ser Met Val Arg Met Lys Ser Met	
80 85 90	
TTT GCT ATT GGC TTT TGT TTT ACT GCC CTA ATG GGA ATG TTC AAT TCC	459
Phe Ala Ile Gly Phe Cys Phe Thr Ala Leu Met Gly Met Phe Asn Ser	
95 100 105	
ATA TTT GAT GGT AGA GTG GTG GCA AAG CTT CCT TTT ACC CCT CTT TCT	507
Ile Phe Asp Gly Arg Val Val Ala Lys Leu Pro Phe Thr Pro Leu Ser	
110 115 120	
TAC ATC CAA GGA CTG TCT CAT CGA AAT CTG CTG GGA GAT GAC ACC ACA	555
Tyr Ile Gln Gly Leu Ser His Arg Asn Leu Leu Gly Asp Asp Thr Thr	
125 130 135	
GAC TGT TCC TTC ATT TTC CTG TAT ATT CTC TGT ACT ATG TCG ATT CGA	603
Asp Cys Ser Phe Ile Phe Leu Tyr Ile Leu Cys Thr Met Ser Ile Arg	
140 145 150 155	
CAG AAC ATT CAG AAG ATT CTC GGC CTT GCC CCT TCA CGA GCC GCC ACC	651
Gln Asn Ile Gln Lys Ile Leu Gly Leu Ala Pro Ser Arg Ala Ala Thr	
160 165 170	
AAG CAG GCA GGT GGA TTT CTT GGC CCA CCA CCT CCT TCT GGG AAG TTC	699
Lys Gln Ala Gly Gly Phe Leu Gly Pro Pro Pro Pro Ser Gly Lys Phe	
175 180 185	
TCT TGA ACTCAAG AACTCTTTAT TTTCTATCAT TCTTTCTAGA CACACACA	750
Ser	
CATCAGACTG GCAACTGTTT TGTAGCAAGA GCCATAGGTA GCCTTACTAC TTGGGCCTCT	810
TTCTAGTTTT GAATTATTTT TAAGCCTTTT GGGTATGATT AGAGTGAAAA TGGCAGCCAG	870
CAAACCTTGAT AGTGCTTTTG GTCCTAGATG ATTTTTATCA AATAAGTGGG TTGATTAGTT	930
AAGTTCAGGT AATGTTTATG TAATGAAAAA CAAATAGCAT CCTTCTTGTT TCATTACAT	990
AAGTATTTTC TGTGGGACCG ACTCTCAAGG CACTGTGTAT GCCCTGCAAG TTGGCTGTCT	1050
ATGAGCATTT AGAGATTTAG AAGAAAAATT TAGTTTGTTC AACCCCTTGTA ACTGTTTGTT	1110
TTGTTGTTGT TTTTTTTTCA AGCCAAATAC ATGACATAAG ATCAATAAAG AGGCCAAATT	1170
TTAGCTGTT TTATGT	1186

Sequence No.: 63

Sequence length: 1409

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Lymphoma

Cell line: U937

Clone name: HP10136

# Sequence characteristics

Code representing characteristics: CDS

Existence site: 82.. 729

Characterization method: E

# Sequence description

```

ATAACTGTTG TCGCGGCGGA GGAAGTGAGG ACGGCGCCAA GGGCCTTCCG GGCCAGTGT      60
GGATCCCTGT AGTTTGTGAA G ATG GTG TTG CTA ACA ATG ATC GCC CGA GTG      111
           Met Val Leu Leu Thr Met Ile Ala Arg Val
           1             5             10
GCG GAC GGG CTC CCG CTG GCC GCC TCG ATG CAG GAG GAC GAA CAG TCT      159
Ala Asp Gly Leu Pro Leu Ala Ala Ser Met Gln Glu Asp Glu Gln Ser
           15             20             25
GGC CGG GAC CTT CAA CAG TAT CAG AGT CAG GCT AAG CAA CTC TTT CGA      207
Gly Arg Asp Leu Gln Gln Tyr Gln Ser Gln Ala Lys Gln Leu Phe Arg
           30             35             40
AAG TTG AAT GAA CAG TCC CCT ACC AGA TGT ACC TTG GAA GCA GGA GCC      255
Lys Leu Asn Glu Gln Ser Pro Thr Arg Cys Thr Leu Glu Ala Gly Ala
           45             50             55
ATG ACT TTT CAC TAC ATT ATT GAG CAG GGG GTG TGT TAT TTG GTT TTA      303
Met Thr Phe His Tyr Ile Ile Glu Gln Gly Val Cys Tyr Leu Val Leu
           60             65             70
TGT GAA GCT GCC TTC CCT AAG AAG TTG GCT TTT GCC TAC CTA GAA GAT      351
Cys Glu Ala Ala Phe Pro Lys Lys Leu Ala Phe Ala Tyr Leu Glu Asp
           75             80             85             90
TTG CAC TCA GAA TTT GAT GAA CAG CAT GGA AAG AAG GTG CCC ACT GTG      399
Leu His Ser Glu Phe Asp Glu Gln His Gly Lys Lys Val Pro Thr Val
           95             100             105
TCC CGA CCC TAT TCC TTT ATT GAA TTT GAT ACT TTC ATT CAG AAA ACC      447
Ser Arg Pro Tyr Ser Phe Ile Glu Phe Asp Thr Phe Ile Gln Lys Thr
           110             115             120
AAG AAG CTC TAC ATT GAC AGT CGT GCT CGA AGA AAT CTA GGC TCC ATC      495
Lys Lys Leu Tyr Ile Asp Ser Arg Ala Arg Arg Asn Leu Gly Ser Ile
           125             130             135
AAC ACT GAA TTG CAA GAT GTG CAG AGG ATC ATG GTG GCC AAT ATT GAA      543
Asn Thr Glu Leu Gln Asp Val Gln Arg Ile Met Val Ala Asn Ile Glu
           140             145             150
GAA GTG TTA CAA CGA GGA GAA GCA CTC TCA GCA TTG GAT TCA AAG GCT      591
Glu Val Leu Gln Arg Gly Glu Ala Leu Ser Ala Leu Asp Ser Lys Ala
           155             160             165             170

```



155

```

AAC AAT TTG TCC AGT CTG TCC AAG AAA TAC CGC CAG GAT GCG AAG TAC      639
Asn Asn Leu Ser Ser Leu Ser Lys Lys Tyr Arg Gln Asp Ala Lys Tyr
      175                      180                      185
TTG AAC ATG CGT TCC ACT TAT GCC AAA CTT GCA GCA GTA GCT GTA TTT      687
Leu Asn Met Arg Ser Thr Tyr Ala Lys Leu Ala Ala Val Ala Val Phe
      190                      195                      200
TTC ATC ATG TTA ATA GTG TAT GTC CGA TTC TGG TGG CTG TGAA      730
Phe Ile Met Leu Ile Val Tyr Val Arg Phe Trp Trp Leu
      205                      210                      215
ATAATGAATA CAGTCACTGG TAAGGGAGAA CCTAGAACCC AGTAGGTGTA TATTTTCAGG      790
AAACTGAGCT CACAGAGATG TGTATTAGAA TCCAAGTGGG ACTTCTGCCT CTAAAGACCT      850
TGCAAGAAAA GAGATGCCCT GAAAATGAAA GGTTGCACCT CATTAAATGA AGCTTAACCC      910
TATGTAGAAA GTCTCTTTTG GGGGCAGAGG CTTTCTCTGG GTGCCAAGCC ATATATATTA      970
GGGAATAGTA GATTGTTAAT TTCGTTTTTT CCCTCCAGT GCATTTTAAA AACAGCACTG     1030
GCTGGGGCAT TCTCATTCTC TGATGGAGCC ATCAATGAGA TTTAACTTAG TCAACCTGTG     1090
CTAGCAACAT TCTGAAATTC CTTCAAAGAA GGCAGTCCTT TGGGAAGGTG TTTTTTTTTT     1150
TTTTTTTTTT TTTGACTCTA ATCAACATTC CTTTGTGTTG TGACATTGTG GATTTTCAGT     1210
AATCTGAGTT TTTGATGGCC TTTTAAACAA GACTCCAGTA TGTGAAGGTT AATTGCTGTG     1270
CTCCACAGAT CTTGTCTATT GGCCCGTGTA GAAAGTTAAC CTTGTGTGTT TTCCTTTTAT     1330
AATTTGCTTA TTGCACAATT GCTTTAGGGT AAGTGAATTA TATTAAGATG CCTTGAAATT     1390
ATAGCACTCC TTGATTAAG                                         1409

```

Sequence No.: 64

Sequence length: 974

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10175

Sequence characteristics

Code representing characteristics: CDS

Existence site: 174.. 512

Characterization method: E

Sequence description

```

AGAGCCGCTC CCCTCTCCTC GCCCGCCAC CGGGACGGAG AGCGCCCGCC GCTGCATTTTC      60
CGGCGACACC TCGCAGTCAT TCCTGCGGCT TGCGCGCCCT TGTAGACAGC CGGGGCCCTTC      120
GTGAGACCGG TGCAGGCCTG GGGTAGTCTC CTGTCTGGAC AGAGAAGAGA AAA ATG      176

```

Met

1

156

CAG GAC ACT GGC TCA GTA GTG CCT TTG CAT TGG TTT GGC TTT GGC TAC	224
Gln Asp Thr Gly Ser Val Val Pro Leu His Trp Phe Gly Phe Gly Tyr	
5 10 15	
GCA GCA CTG GTT GCT TCT GGT GGG ATC ATT GGC TAT GTA AAA GCA GGC	272
Ala Ala Leu Val Ala Ser Gly Gly Ile Ile Gly Tyr Val Lys Ala Gly	
20 25 30	
AGC GTG CCG TCC CTG GCT GCA GGG CTG CTC TTT GGC AGT CTA GCC GGC	320
Ser Val Pro Ser Leu Ala Ala Gly Leu Leu Phe Gly Ser Leu Ala Gly	
35 40 45	
CTG GGT GCT TAC CAG CTG TCT CAG GAT CCA AGG AAC GTT TGG GTT TTC	368
Leu Gly Ala Tyr Gln Leu Ser Gln Asp Pro Arg Asn Val Trp Val Phe	
50 55 60 65	
CTA GCT ACA TCT GGT ACC TTG GCT GGC ATT ATG GGA ATG AGG TTC TAC	416
Leu Ala Thr Ser Gly Thr Leu Ala Gly Ile Met Gly Met Arg Phe Tyr	
70 75 80	
CAC TCT GGA AAA TTC ATG CCT GCA GGT TTA ATT GCA GGT GCC AGT TTG	464
His Ser Gly Lys Phe Met Pro Ala Gly Leu Ile Ala Gly Ala Ser Leu	
85 90 95	
CTG ATG GTC GCC AAA GTT GGA GTT AGT ATG TTC AAC AGA CCC CAT	509
Leu Met Val Ala Lys Val Gly Val Ser Met Phe Asn Arg Pro His	
100 105 110	
T AGCAGAACTC ATGTTCCAGC TTAGACTGAT GAAGAATTAA AAATCTGCAT	560
CTTCCACTAT TTTCAATATA TTAAGAGAAA TAAGTGCAGC ATTTTTCAT CTGACATTTT	620
ACCTAAAAAA AAAGACACCA AACTTGGCAG AGAGGTGGAA AATCAGTCAT GATTACAAAC	680
CTACAGAGGT GCGGAGTATG TAACACAAGA GCTTAATAAG ACCCTCATAG AGCTTGATTC	740
TTGTATATTG ATGTTGTCTT TTCTTTCTGT ATCTGTAGGT AAATCTCAAG GGTAATAATGT	800
TAGGTGTCAG CTTTCAGGGC TCTGAAACCC TATTCCTGCT TCTGAGGAAC AGTGTGAAAA	860
AAAGTCTTTT AGGAGATTTA CAATATCTGT TCTTTTGCTC ATCTTAGACC ACAGACTGAC	920
TTTGAAATTA TGTTAAGTGA AATATCAATG TAAATAAAGT TTACTATAAA TAAT	974

Sequence No.: 65

Sequence length: 925

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP10179

Sequence characteristics

Code representing characteristics: CDS

Existence site: 122.. 466

Characterization method: E

Sequence description

```

AATCGCGTTT CCGGAGAGAC CTGGCTGCTG TGTCCCGCGG CTTGCGCTCC GTAGTGGACT      60
CCGCGGGCCT TCGGCAGATG CAGGCCTGGG GTAGTCTCCT TTCTGGACTG AGAAGAGAAG      120
ATG GAG AAG CCC CTC TTC CCA TTA GTG CCT TTG CAT TGG TTT GGC TTT      168
Met Glu Lys Pro Leu Phe Pro Leu Val Pro Leu His Trp Phe Gly Phe
   1             5             10            15
GGC TAC ACA GCA CTG GTT GTT TCT GGT GGG ATC GTT GGC TAT GTA AAA      216
Gly Tyr Thr Ala Leu Val Val Ser Gly Gly Ile Val Gly Tyr Val Lys
           20           25           30
ACA GGC AGC GTG CCG TCC CTG GCA GCA GGG CTG CTC TTC GGC AGT CTA      264
Thr Gly Ser Val Pro Ser Leu Ala Ala Gly Leu Leu Phe Gly Ser Leu
           35           40           45
GCC GGC CTG GGT GCT TAC CAG CTG TAT CAG GAT CCT AGG AAC GTT TGG      312
Ala Gly Leu Gly Ala Tyr Gln Leu Tyr Gln Asp Pro Arg Asn Val Trp
           50           55           60
GGT TTC CTA GCC GCT ACA TCT GTT ACT TTT GTT GGT GTT ATG GGA ATA      360
Gly Phe Leu Ala Ala Thr Ser Val Thr Phe Val Gly Val Met Gly Met
           65           70           75           80
AGA TCC TAC TAC TAT GGA AAA TTC ATG CCT GTA GGT TTA ATT GCA GGT      408
Arg Ser Tyr Tyr Tyr Gly Lys Phe Met Pro Val Gly Leu Ile Ala Gly
           85           90           95
GCC AGT TTG CTG ATG GCC GCC AAA GTT GGA GTT CGT ATG TTG ATG ACA      456
Ala Ser Leu Leu Met Ala Ala Lys Val Gly Val Arg Met Leu Met Thr
           100          105          110
TCT GAT TAGCAGAAGT CATGTTGCA GCTTGGACTC ATGAAGGATT AAAAATCT      510
Ser Asp

GCATCTTCCA CTATTTTCAA TGTATTAAGA GAAATAAGTG CAGCATTTTT GCATCTGACA      570
TTTTACCTAA AAAAAAAAAAG ACACCAAATT TGGCGGAGGG GTGAAAAATC AGTTGTTACC      630
ATTATAACCC TACAGAGGTG GTGAGCATGT AACATGAGCT TATTGAGACC ATCATAGAGA      690
TCGATTCTTG TATATTGATT TTATCTCTTT CTGTATCTAT AGGTAAATCT CAAGGGTAAA      750
ATGTTAGGTG TTGACATTGA GAACCCTGAA ACCCCATTCC CTGCTCAGAG GAACAGTGTG      810
AAAAAAAATC TCTTGAGAGA TTTAGAATAT CTTTTCTTTT GCTCATCTTA GACCACAGAC      870
TGACTTTGAA ATTATGTTAA GTGAAATATC AATGAAAATA AAGTTTACTA TAAAT      925

```

Sequence No.: 66

Sequence length: 1115

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Fibrosarcoma

Cell line: HT-1080

Clone name: HP10196

Sequence characteristics

Code representing characteristics: CDS

Existence site: 10.. 993

Characterization method: E

Sequence description

CGCGGGGAAA ATG GCG GCG GCG GCG GCG GCG GCT GCA GCT ACG AAC GGG ACC	51
Met Ala Ala Ala Ala Ala Ala Ala Ala Ala Thr Asn Gly Thr	
1 5 10	
GGA GGA AGC AGC GGG ATG GAG GTG GAT GCA GCA GTA GTC CCC AGC GTG	99
Gly Gly Ser Ser Gly Met Glu Val Asp Ala Ala Val Val Pro Ser Val	
15 20 25 30	
ATG GCC TGC GGA GTG ACT GGG AGT GTT TCC GTC GCT CTC CAT CCC CTT	147
Met Ala Cys Gly Val Thr Gly Ser Val Ser Val Ala Leu His Pro Leu	
35 40 45	
GTC ATT CTC AAC ATC TCA GAC CAC TGG ATC CGC ATG CGC TCC CAG GAG	195
Val Ile Leu Asn Ile Ser Asp His Trp Ile Arg Met Arg Ser Gln Glu	
50 55 60	
GGG CGG CCT GTG CAG GTG ATT GCG GCT CTG ATT GGC AAG CAG GAG GGC	243
Gly Arg Pro Val Gln Val Ile Gly Ala Leu Ile Gly Lys Gln Glu Gly	
65 70 75	
CGA AAT ATC GAG GTG ATG AAC TCC TTT GAG CTG CTG TCC CAC ACC GTG	291
Arg Asn Ile Glu Val Met Asn Ser Phe Glu Leu Leu Ser His Thr Val	
80 85 90	
GAA GAG AAG ATT ATC ATT CAC AAG GAA TAT TAT TAC ACC AAG GAG GAG	339
Glu Glu Lys Ile Ile Ile Asp Lys Glu Tyr Tyr Tyr Thr Lys Glu Glu	
95 100 105 110	
CAG TTT AAA CAG GTG TTC AAG GAG CTG GAG TTT CTG GGT TGG TAT ACC	387
Gln Phe Lys Gln Val Phe Lys Glu Leu Glu Phe Leu Gly Trp Tyr Thr	
115 120 125	
ACA GCG GGG CCA CCT GAC CCC TCG GAC ATC CAC GTC CAT AAG CAG GTG	435
Thr Gly Gly Pro Pro Asp Pro Ser Asp Ile His Val His Lys Gln Val	
130 135 140	
TGT GAG ATC ATC GAG AGC CCC CTC TTT CTG AAG TTG AAC CCT ATG ACC	483
Cys Glu Ile Ile Glu Ser Pro Leu Phe Leu Lys Leu Asn Pro Met Thr	
145 150 155	
AAG CAC ACA GAT CTT CCT GTC AGC GTT TTT GAG TCT GTC ATT GAT ATA	531
Lys His Thr Asp Leu Pro Val Ser Val Phe Glu Ser Val Ile Asp Ile	

160	165	170	
ATC AAT GGA GAG GCC ACA ATG CTG TTT GCT GAG CTG ACC TAC ACT CTG			579
Ile Asn Gly Glu Ala Thr Met Leu Phe Ala Glu Leu Thr Tyr Thr Leu			
175	180	185	190
GCC ACA GAG GAA GCG GAA CGC ATT GGT GTA GAC CAC GTA GCC CGA ATG			627
Ala Thr Glu Glu Ala Glu Arg Ile Gly Val Asp His Val Ala Arg Met			
	195	200	205
ACA GCA ACA GGC AGT GGA GAG AAC TCC ACT GTG GCT GAA CAC CTG ATA			675
Thr Ala Thr Gly Ser Gly Glu Asn Ser Thr Val Ala Glu His Leu Ile			
	210	215	220
GCA CAG CAC AGC GCC ATC AAG ATG CTG CAC AGC CGC GTC AAG CTC ATC			723
Ala Gln His Ser Ala Ile Lys Met Leu His Ser Arg Val Lys Leu Ile			
	225	230	235
TTG GAG TAC GTC AAG GCC TCT GAA GCG GGA GAG GTC CCC TTT AAT CAT			771
Leu Glu Tyr Val Lys Ala Ser Glu Ala Gly Glu Val Pro Phe Asn His			
	240	245	250
GAG ATC CTG CGG GAG GCC TAT GCT CTG TGT CAC TGT CTC CCG GTG CTC			819
Glu Ile Leu Arg Glu Ala Tyr Ala Leu Cys His Cys Leu Pro Val Leu			
	255	260	265
AGC ACA GAC AAG TTC AAG ACA GAT TTT TAT GAT CAA TGC AAC GAC GTG			867
Ser Thr Asp Lys Phe Lys Thr Asp Phe Tyr Asp Gln Cys Asn Asp Val			
	275	280	285
GGG CTC ATG GCC TAC CTC GGC ACC ATC ACC AAA ACG TGC AAC ACC ATG			915
Gly Leu Met Ala Tyr Leu Gly Thr Ile Thr Lys Thr Cys Asn Thr Met			
	290	295	300
AAC CAG TTT GTG AAC AAG TTC AAT GTC CTC TAC GAC CGA CAA GGC ATC			963
Asn Gln Phe Val Asn Lys Phe Asn Val Leu Tyr Asp Arg Gln Gly Ile			
	305	310	315
GGC AGG AGA ATG CGC GGC CTC TTT TTC TGATGAGGGT			1000
Gly Arg Arg Met Arg Gly Leu Phe Phe			
	320	325	
ACTTGAAGGG CTGATGGACA GGGGTCAGGC AACTATCCCA AAGGGGAGGG CACTACACTT			1060
CCTTGAGAGA AACCACTGTC ATTAATAAAA GGGGAGCAGC CCCTGAGCAC CCCTG			1115

Sequence No.: 67

Sequence length: 1721

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Fibrosarcoma

Cell line: HT-1080

Clone name: HP10235

## Sequence characteristics

Code representing characteristics: CDS

Existence site: 6.. 1127

Characterization method: E

## Sequence description

ATGTC ATG ACC CTA TGT GCC ATG CTG CCC CTG CTG TTA TTC ACC TAC CTC	50
Met Thr Leu Cys Ala Met Leu Pro Leu Leu Phe Thr Tyr Leu	
1 5 10 15	
AAC TCC TTC CTG CAT CAG AGG ATC CCC CAG TCC GTA CCG ATC CTG GGC	98
Asn Ser Phe Leu His Gln Arg Ile Pro Gln Ser Val Arg Ile Leu Gly	
20 25 30	
AGC CTG GTG GCC ATC CTG CTG GTG TTT CTG ATC ACT GCC ATC CTG GTG	146
Ser Leu Val Ala Ile Leu Leu Val Phe Leu Ile Thr Ala Ile Leu Val	
35 40 45	
AAG GTG CAG CTG GAT GCT CTG CCC TTC TTT GTC ATC ACC ATG ATC AAG	194
Lys Val Gln Leu Asp Ala Leu Pro Phe Phe Val Ile Thr Met Ile Lys	
50 55 60	
ATC GTG CTC ATT AAT TCA TTT GGT GCC ATC CTG CAG GGC AGC CTG TTT	242
Ile Val Leu Ile Asn Ser Phe Gly Ala Ile Leu Gln Gly Ser Leu Phe	
65 70 75	
GGT CTG GCT GGC CTT CTG CCT GCC AGC TAC ACG GCC CCC ATC ATG AGT	290
Gly Leu Ala Gly Leu Leu Pro Ala Ser Tyr Thr Ala Pro Ile Met Ser	
80 85 90 95	
GGC CAG GGC CTA GCA GGC TTC TTT GCC TCC GTG GCC ATG ATC TGC GCT	338
Gly Gln Gly Leu Ala Gly Phe Phe Ala Ser Val Ala Met Ile Cys Ala	
100 105 110	
ATT GCC AGT GGC TCG GAG CTA TCA GAA AGT GCC TTC GGC TAC TTT ATC	386
Ile Ala Ser Gly Ser Glu Leu Ser Glu Ser Ala Phe Gly Tyr Phe Ile	
115 120 125	
ACA GCC TGT GCT GTT ATC ATT TTG ACC ATC ATC TGT TAC CTG GGC CTG	434
Thr Ala Cys Ala Val Ile Ile Leu Thr Ile Ile Cys Tyr Leu Gly Leu	
130 135 140	
CCC CGC CTG GAA TTC TAC CGC TAC TAC CAG CAG CTC AAG CTT GAA GGA	482
Pro Arg Leu Glu Phe Tyr Arg Tyr Tyr Gln Gln Leu Lys Leu Glu Gly	
145 150 155	
CCC GGG GAG CAG GAG ACC AAG TTG GAC CTC ATT AGC AAA GGA GAG GAG	530
Pro Gly Glu Gln Glu Thr Lys Leu Asp Leu Ile Ser Lys Gly Glu Glu	
160 165 170 175	
CCA AGA GCA GGC AAA GAG GAA TCT GGA GTT TCA GTC TCC AAC TCT CAG	578
Pro Arg Ala Gly Lys Glu Glu Ser Gly Val Ser Val Ser Asn Ser Gln	
180 185 190	

CCC ACC AAT GAA AGC CAC TCT ATC AAA GCC ATC CTG AAA AAT ATC TCA	626
Pro Thr Asn Glu Ser His Ser Ile Lys Ala Ile Leu Lys Asn Ile Ser	
195 200 205	
GTC CTG GCT TTC TCT GTC TGC TTC ATC TTC ACT ATC ACC ATT GGG ATG	674
Val Leu Ala Phe Ser Val Cys Phe Ile Phe Thr Ile Thr Ile Gly Met	
210 215 220	
TTT CCA GCC GTG ACT GTT GAG GTC AAG TCC AGC ATC GCA GGC AGC AGC	722
Phe Pro Ala Val Thr Val Glu Val Lys Ser Ser Ile Ala Gly Ser Ser	
225 230 235	
ACC TGG GAA CGT TAC TTC ATT CCT GTG TCC TGT TTC TTG ACT TTC AAT	770
Thr Trp Glu Arg Tyr Phe Ile Pro Val Ser Cys Phe Leu Thr Phe Asn	
240 245 250 255	
ATC TTT GAC TGG TTG GGC CGG AGC CTC ACA GCT GTA TTC ATG TGG CCT	818
Ile Phe Asp Trp Leu Gly Arg Ser Leu Thr Ala Val Phe Met Trp Pro	
260 265 270	
GGG AAG GAC AGC CGC TGG CTG CCA AGC CTG GTC CTG GCC CGG CTG GTG	866
Gly Lys Asp Ser Arg Trp Leu Pro Ser Leu Val Leu Ala Arg Leu Val	
275 280 285	
TTT GTG CCA CTG CTG CTG CTG TGC AAC ATT AAG CCC CGC CGC TAC CTG	914
Phe Val Pro Leu Leu Leu Leu Cys Asn Ile Lys Pro Arg Arg Tyr Leu	
290 295 300	
ACT GTG GTC TTC GAG CAC GAT GCC TGG TTC ATC TTC TTC ATG GCT GCC	962
Thr Val Val Phe Glu His Asp Ala Trp Phe Ile Phe Phe Met Ala Ala	
305 310 315	
TTT GCC TTC TCC AAC GGC TAC CTC GCC AGC CTC TGC ATG TGC TTC GGG	1010
Phe Ala Phe Ser Asn Gly Tyr Leu Ala Ser Leu Cys Met Cys Phe Gly	
320 325 330 335	
CCC AAG AAA GTG AAG CCA GCT GAG GCA GAG ACC GCA GGA GCC ATC ATG	1058
Pro Lys Lys Val Lys Pro Ala Glu Ala Glu Thr Ala Gly Ala Ile Met	
340 345 350	
GCC TTC TTC CTG TGT CTG GGT CTG GCA CTG GGG GCT GTT TTC TCC TTC	1106
Ala Phe Phe Leu Cys Leu Gly Leu Ala Leu Gly Ala Val Phe Ser Phe	
355 360 365	
CTG TTC CGG GCA ATT GTG TGACAAAGGA TGGACAGAAG GACTGC	1150
Leu Phe Arg Ala Ile Val	
370	
CTGCCTCCCT CCCTGTCTGC CTCCTGCCCC TTCCTTCTGC CAGGGGTGAT CCTGAGTGGT	1210
CTGGCGGTTT TTTCTTCTAA CTGACTTCTG CTTTCCACGG CGTGTGCTGG GCGCGGATCT	1270
CCAGGCCCTG GGGAGGGAGC CTCTGGACGG ACAGTGGGGA CATTGTGGGT TTGGGGCTCA	1330
GAGTCGAGGG ACGGGGTGTA GCCTCGGCAT TTGCTTGAGT TTCTCCACTC TTGGCTCTGA	1390
CTGATCCCTG CTTGTGCAGG CCAGTGGAGG CTCTTGGGCT TGGAGAACAC GTGTGTCTCT	1450
GTGTATGTGT CTGTGTGTCT GCGTCCGTGT CTGTGCACT GTCTGCCCTGT CCTGGGGTGG	1510
CTAGGAGCTG GGTCTGACCG TTGTATGGTT TGACCTGATA TACTCCATTC TCCCCTGCGC	1570
CTCCTCCTCT GTGTTCTCTC CATGTCCCC TCCCAACTCC CCATGCCAG TTCTTACCCA	1630

TCATGCACCC TGTACAGTTG CCACGTTACT GCCTTTTTTA AAAATATATT TGACAGAAAC 1690  
 CAGGTGCCTT CAGAGGCTCT CTGATTTAAA T 1721

Sequence No.: 68

Sequence length: 1504

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10297

Sequence characteristics

Code representing characteristics: CDS

Existence site: 63.. 614

Characterization method: E

Sequence description

CTTTTGC GGC TGCAGCGGGC TTGTAGGTGT CCGGCTTTGC TGGCCCAGCA AGCCTGATAA 60  
 GC ATG AAG CTC TTA TCT TTG GTG GCT GTG GTC GGG TGT TTG CTG GTG 107  
 Met Lys Leu Leu Ser Leu Val Ala Val Val Gly Cys Leu Leu Val  
 1 5 10 15  
 CCC CCA GCT GAA GCC AAC AAG AGT TCT GAA GAT ATC CGG TGC AAA TGC 155  
 Pro Pro Ala Glu Ala Asn Lys Ser Ser Glu Asp Ile Arg Cys Lys Cys  
 20 25 30  
 ATC TGT CCA CCT TAT AGA AAC ATC AGT GGG CAC ATT TAC AAC CAG AAT 203  
 Ile Cys Pro Pro Tyr Arg Asn Ile Ser Gly His Ile Tyr Asn Gln Asn  
 35 40 45  
 GTA TCC CAG AAG GAC TGC AAC TGC CTG CAC GTG GTG GAG CCC ATG CCA 251  
 Val Ser Gln Lys Asp Cys Asn Cys Leu His Val Val Glu Pro Met Pro  
 50 55 60  
 GTG CCT GGC CAT GAC GTG GAG GCC TAC TGC CTG CTG TGC GAG TGC AGG 299  
 Val Pro Gly His Asp Val Glu Ala Tyr Cys Leu Leu Cys Glu Cys Arg  
 65 70 75  
 TAC GAG GAG CGC AGC ACC ACC ACC ATC AAG GTC ATC ATT GTC ATC TAC 347  
 Tyr Glu Glu Arg Ser Thr Thr Thr Ile Lys Val Ile Ile Val Ile Tyr  
 80 85 90 95  
 CTG TCC GTG GTG GGT GCC CTG TTG CTC TAC ATG GCC TTC CTG ATG CTG 395  
 Leu Ser Val Val Gly Ala Leu Leu Leu Tyr Met Ala Phe Leu Met Leu  
 100 105 110  
 GTG GAC CCT CTG ATC CGA AAG CCG GAT GCA TAC ACT GAG CAA CTG CAC 443  
 Val Asp Pro Leu Ile Arg Lys Pro Asp Ala Tyr Thr Glu Gln Leu His



163

115	120	125	
AAT GAG GAG GAG AAT GAG GAT GCT CGC TCT ATG GCA GCA GCT GCT GCA			491
Asn Glu Glu Glu Asn Glu Asp Ala Arg Ser Met Ala Ala Ala Ala Ala			
130	135	140	
TCC CTC GGG GGA CCC CGA GCA AAC ACA GTC CTG GAG CGT GTG GAA GGT			539
Ser Leu Gly Gly Pro Arg Ala Asn Thr Val Leu Glu Arg Val Glu Gly			
145	150	155	
GCC CAG CAG CGG TGG AAG CTG CAG GTG CAG GAG CAG CGG AAG ACA GTC			587
Ala Gln Gln Arg Trp Lys Leu Gln Val Gln Glu Gln Arg Lys Thr Val			
160	165	170	175
TTC GAT CGG CAC AAG ATG CTC AGC TAGATGGGCT GGTGTGGTTG GGTCAAGGC			640
Phe Asp Arg His Lys Met Leu Ser			
180			
CCCAACACCA TGGCTGCCAG CTTCCAGGCT GGACAAAGCA GGGGGCTACT TCTCCCTTCC			700
CTCGGTTCCA GTCTTCCCTT TAAAAGCCTG TGGCATTITT CCTCCTTCTC CCTAACTTTA			760
GAAATGTTGT ACTTGCTAT TTTGATTAGG GAAGAGGGAT GTGGTCTCTG ATCTCTGTTG			820
TCTTCTTGGG TCTTTGGGGT TGAAGGGAGG GCGAAGGCAG GCCAGAAGCG AATGGAGACA			880
TTGAGGCGG CCTCAGGAGT GGATGCGATC TGTCTCTCCT GGCTCCACTC TTGCCGCCTT			940
CCAGCTCTGA GTCTTGGGAA TGTGTTACC CTTGGAAGAT AAAGCTGGGT CTTCAAGAAC			1000
TCAGTGTCTG GGAGGAAAGC ATGGCCAGC ATTCAAGCATG TGTTCCTTTC TGCAGTGGTT			1060
CTTATCACCA CCTCCCTCCC AGCCCCAGCG CCTCAGCCCC AGCCCCAGCT CCAGCCCTGA			1120
GGACAGCTCT GATGGGAGAG CTGGGCCCCC TGAGCCCACT GGGTCTTCAG GGTGCACTGG			1180
AAGCTGGTGT TCGCTGTCCC CTGTGCACTT CTCGCACTGG GGCATGGAGT GCCCATGCAT			1240
ACTCTGCTGC CGGTCCCCTC ACCTGCACTT GAGGGGTCTG GGCAGTCCCT CCTCTCCCCA			1300
GTGTCCACAG TCACTGAGCC AGACGGTCGG TTGGAACATG AGACTCGAGG CTGAGCGTGG			1360
ATCTGAACAC CACAGCCCCT GTACTTGGGT TGCCTCTTGT CCCTGAACCT CGTTGTACCA			1420
GTGCATGGAG AGAAAATTTT GTCCTCTTGT CTTAGAGTTG TGTGTAAATC AAGGAAGCCA			1480
TCATTAAATT GTTTTATTTT TCTC			1504

Sequence No.: 69

Sequence length: 532

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Stomach cancer

Clone name: HP10299

Sequence characteristics

Code representing characteristics: CDS

Existence site: 93.. 443

Characterization method: E

## Sequence description

```

GCTCTCTGGT AAAGGCGTGC AGGTGTTGGC CGCGGCCTCT GAGCTGGGAT GAGCCGTGCT      60
CCCCGTGCAA GCAAGGGAGC CCAGCCGGAG CC ATG GCC AGT ACA GTG GTA GCA      113
                               Met Ala Ser Thr Val Val Ala
                               1           5

GTT GGA CTG ACC ATT GCT GCT GCA GGA TTT GCA GGC CGT TAC GTT TTG      161
Val Gly Leu Thr Ile Ala Ala Ala Gly Phe Ala Gly Arg Tyr Val Leu
      10           15           20

CAA GCC ATG AAG CAT ATG GAG CCT CAA GTA AAA CAA GTT TTT CAA AGC      209
Gln Ala Met Lys His Met Glu Pro Gln Val Lys Gln Val Phe Gln Ser
      25           30           35

CTA CCA AAA TCT GCC TTC AGT GGT GGC TAT TAT AGA GGT GGG TTT GAA      257
Leu Pro Lys Ser Ala Phe Ser Gly Gly Tyr Tyr Arg Gly Gly Phe Glu
      40           45           50           55

CCC AAA ATG ACA AAA CGG GAA GCA GCA TTA ATA CTA GGT GTA AGC CCT      305
Pro Lys Met Thr Lys Arg Glu Ala Ala Leu Ile Leu Gly Val Ser Pro
      60           65           70

ACT GCC AAT AAA GGG AAA ATA AGA GAT GCT CAT CGA CGA ATT ATG CTT      353
Thr Ala Asn Lys Gly Lys Ile Arg Asp Ala His Arg Arg Ile Met Leu
      75           80           85

TTA AAT CAT CCT GAC AAA GGA GGA TCT CCT TAT ATA GCA GCC AAA ATC      401
Leu Asn His Pro Asp Lys Gly Gly Ser Pro Tyr Ile Ala Ala Lys Ile
      90           95           100

AAT GAA GCT AAA GAT TTA CTA GAA GGT CAA GCT AAA AAA TGAAGTAAAT      450
Asn Glu Ala Lys Asp Leu Leu Glu Gly Gln Ala Lys Lys
      105           110           115

GTATGATGAA TTTTAAGTTC GTATTAGTTT ATGTATATGA GTACTAAGTT TTTATAATAA      510
AATGCCTCAG AGCTACAATT TT      532

```

Sequence No.: 70

Sequence length: 662

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP10301

Sequence characteristics

Code representing characteristics: CDS

Existence site: 92.. 550

Characterization method: E

Sequence description

```

TCTAGCCCCG CCCCAGGCGA GGGCGCCGCA CCCACACCGC GCTGCGCAGT TTTGTTCTGC      60
TCCAGCTGTT CGAAGGTGAT CCAGACGCAA G ATG GCT GTC CTC TCT AAG GAA      112
                               Met Ala Val Leu Ser Lys Glu
                               1           5
TAT GGT TTT GTG CTT CTA ACT GGT GCT GCC AGC TTT ATA ATG GTG GCC      160
Tyr Gly Phe Val Leu Leu Thr Gly Ala Ala Ser Phe Ile Met Val Ala
      10           15           20
CAC CTA GCC ATC AAT GTT TCC AAG GCC CGC AAG AAG TAC AAA GTG GAG      208
His Leu Ala Ile Asn Val Ser Lys Ala Arg Lys Lys Tyr Lys Val Glu
      25           30           35
TAT CCT ATC ATG TAC AGC ACG GAC CCT GAA AAT GGG CAC ATC TTC AAC      256
Tyr Pro Ile Met Tyr Ser Thr Asp Pro Glu Asn Gly His Ile Phe Asn
      40           45           50           55
TGC ATT CAG CGA GCC CAC CAG AAC ACG TTG GAA GTG TAT CCT CCC TTC      304
Cys Ile Gln Arg Ala His Gln Asn Thr Leu Glu Val Tyr Pro Pro Phe
      60           65           70
TTA TTT TTT CTA GCT GTT GGA GGT GTT TAC CAC CCG CGT ATA GCT TCT      352
Leu Phe Phe Leu Ala Val Gly Gly Val Tyr His Pro Arg Ile Ala Ser
      75           80           85
GGC CTG GGC TTG GCC TGG ATT GTT GGA CGA GTT CTT TAT GCT TAT GGC      400
Gly Leu Gly Leu Ala Trp Ile Val Gly Arg Val Leu Tyr Ala Tyr Gly
      90           95           100
TAT TAC ACG GGA GAA CCC AGC AAG CGT AGT CGA GGA GCC CTG GGG TCC      448
Tyr Tyr Thr Gly Glu Pro Ser Lys Arg Ser Arg Gly Ala Leu Gly Ser
      105           110           115
ATC GCC CTC CTG GGC TTG GTG GGC ACA ACT GTG TGC TCT GCT TTC CAG      496
Ile Ala Leu Leu Gly Leu Val Gly Thr Thr Val Cys Ser Ala Phe Gln
      120           125           130           135
CAT CTT GGT TGG GTT AAA AGT GGC TTG GGC AGT GGA CCC AAA TGC TGC      544
His Leu Gly Trp Val Lys Ser Gly Leu Gly Ser Gly Pro Lys Cys Cys
      140           145           150
CAT TAAAGAATTA TAGGGGTTTA AAAACTCTCA TTCATTTTAA ATG      590
His

ACTTACCTTT ATTTCCAGTT ACATTTTTTT TCTAAATATA ATAAAAACTT ACCTGGCATC      650
AGCCTCATAC CT      662

```

Sequence No.: 71

Sequence length: 2373

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Liver

Clone name: HP10302

Sequence characteristics

Code representing characteristics: CDS

Existence site: 134.. 1813

Characterization method: E

Sequence description

```

GAAGACCCCA GCGCCGGCGC GGCTCAGGGC TGGGCCACG GGA CTCCGGA CGCGCCGCGA      60
AAGCGTTGCG CTCCCGGAGG CGTCCGCAGC TGCTGGCTGC TCATTGCCC GTGACCGGAG      120
GCTCGGGGCC AGC ATG GCC CCC ACG CTG CAA CAG GCG TAC CGG AGG CGC      169
      Met Ala Pro Thr Leu Gln Gln Ala Tyr Arg Arg Arg
              1              5              10
TGG TGG ATG GCC TGC ACG GCT GTG CTG GAG AAC CTC TTC TTC TCT GCT      217
Trp Trp Met Ala Cys Thr Ala Val Leu Glu Asn Leu Phe Phe Ser Ala
              15              20              25
GTA CTC CTG GGC TGG GGC TCC CTG TTG ATC ATT CTG AAG AAC GAG GGC      265
Val Leu Leu Gly Trp Gly Ser Leu Leu Ile Ile Leu Lys Asn Glu Gly
              30              35              40
TTC TAT TCC AGC ACG TGC CCA GCT GAG AGC AGC ACC AAC ACC ACC CAG      313
Phe Tyr Ser Ser Thr Cys Pro Ala Glu Ser Ser Thr Asn Thr Thr Gln
              45              50              55              60
GAT GAG CAG CGC AGG TGG CCA GGC TGT GAC CAG CAG GAC GAG ATG CTC      361
Asp Glu Gln Arg Arg Trp Pro Gly Cys Asp Gln Gln Asp Glu Met Leu
              65              70              75
AAC CTG GGC TTC ACC ATT GGT TCC TTC GTG CTC AGC GCC ACC ACC CTG      409
Asn Leu Gly Phe Thr Ile Gly Ser Phe Val Leu Ser Ala Thr Thr Leu
              80              85              90
CCA CTG GGG ATC CTC ATG GAC CGC TTT GGC CCC CGA CCC GTG CGG CTG      457
Pro Leu Gly Ile Leu Met Asp Arg Phe Gly Pro Arg Pro Val Arg Leu
              95              100              105
GTT GGC AGT GCC TGC TTC ACT GCG TCC TGC ACC CTC ATG GCC CTG GCC      505
Val Gly Ser Ala Cys Phe Thr Ala Ser Cys Thr Leu Met Ala Leu Ala
              110              115              120
TCC CGG GAC GTG GAA GCT CTG TCT CCG TTG ATA TTC CTG GCG CTG TCC      553
Ser Arg Asp Val Glu Ala Leu Ser Pro Leu Ile Phe Leu Ala Leu Ser
125              130              135              140

```

167

CTG AAT GGC TTT GGT GGC ATC TGC CTA ACG TTC ACT TCA CTC ACG CTG	601
Leu Asn Gly Phe Gly Gly Ile Cys Leu Thr Phe Thr Ser Leu Thr Leu	
145 150 155	
CCC AAC ATG TTT GGG AAC CTG CGC TCC ACG TTA ATG GCC CTC ATG ATT	649
Pro Asn Met Phe Gly Asn Leu Arg Ser Thr Leu Met Ala Leu Met Ile	
160 165 170	
GGC TCT TAC GCC TCT TCT GCC ATT ACG TTC CCA GGA ATC AAG CTG ATC	697
Gly Ser Tyr Ala Ser Ser Ala Ile Thr Phe Pro Gly Ile Lys Leu Ile	
175 180 185	
TAC GAT GCC GGT GTG GCC TTC GTG GTC ATC ATG TTC ACC TGG TCT GGC	745
Tyr Asp Ala Gly Val Ala Phe Val Val Ile Met Phe Thr Trp Ser Gly	
190 195 200	
CTG GCC TGC CTT ATC TTT CTG AAC TGC ACC CTC AAC TGG CCC ATC GAA	793
Leu Ala Cys Leu Ile Phe Leu Asn Cys Thr Leu Asn Trp Pro Ile Glu	
205 210 215 220	
GCC TTT CCT GCC CCT GAG GAA GTC AAT TAC ACG AAG AAG ATC AAG CTG	841
Ala Phe Pro Ala Pro Glu Glu Val Asn Tyr Thr Lys Lys Ile Lys Leu	
225 230 235	
AGT GGG CTG GCC CTG GAC CAC AAG GTG ACA GGT GAC CTC TTC TAC ACC	889
Ser Gly Leu Ala Leu Asp His Lys Val Thr Gly Asp Leu Phe Tyr Thr	
240 245 250	
CAT GTG ACC ACC ATG GGC CAG AGG CTC AGC CAG AAG GCC CCC AGC CTG	937
His Val Thr Thr Met Gly Gln Arg Leu Ser Gln Lys Ala Pro Ser Leu	
255 260 265	
GAG GAC GGT TCG GAT GCC TTC ATG TCA CCC CAG GAT GTT CGG GGC ACC	985
Glu Asp Gly Ser Asp Ala Phe Met Ser Pro Gln Asp Val Arg Gly Thr	
270 275 280	
TCA GAA AAC CTT CCT GAG AGG TCT GTC CCC TTA CGC AAG AGC CTC TGC	1033
Ser Glu Asn Leu Pro Glu Arg Ser Val Pro Leu Arg Lys Ser Leu Cys	
285 290 295 300	
TCC CCC ACT TTC CTG TGG AGC CTC CTC ACC ATG GGC ATG ACC CAG CTG	1081
Ser Pro Thr Phe Leu Trp Ser Leu Leu Thr Met Gly Met Thr Gln Leu	
305 310 315	
CGG ATC ATC TTC TAC ATG GCT GCT GTG AAC AAG ATG CTG GAG TAC CTT	1129
Arg Ile Ile Phe Tyr Met Ala Ala Val Asn Lys Met Leu Glu Tyr Leu	
320 325 330	
GTG ACT GGT GGC CAG GAG CAT GAG ACA AAT GAA CAG CAA CAA AAG GTG	1177
Val Thr Gly Gly Gln Glu His Glu Thr Asn Glu Gln Gln Gln Lys Val	
335 340 345	
GCA GAG ACA GTT GGG TTC TAC TCC TCC GTC TTC GGG GCC ATG CAG CTG	1225
Ala Glu Thr Val Gly Phe Tyr Ser Ser Val Phe Gly Ala Met Gln Leu	
350 355 360	
TTG TGC CTT CTC ACC TGC CCC CTC ATT GGC TAC ATC ATG GAC TGG CGG	1273
Leu Cys Leu Leu Thr Cys Pro Leu Ile Gly Tyr Ile Met Asp Trp Arg	

365	370	375	380	
ATC AAG GAC TGC GTG GAC GCC CCA ACT CAG GGC ACT GTC CTC GGA GAT				1321
Ile Lys Asp Cys Val Asp Ala Pro Thr Gln Gly Thr Val Leu Gly Asp				
	385	390	395	
GCC AGG GAC GGG GTT GCT ACC AAA TCC ATC AGA CCA CGC TAC TGC AAG				1369
Ala Arg Asp Gly Val Ala Thr Lys Ser Ile Arg Pro Arg Tyr Cys Lys				
	400	405	410	
ATC CAA AAG CTC ACC AAT GCC ATC AGT GCC TTC ACC CTG ACC AAC CTG				1417
Ile Gln Lys Leu Thr Asn Ala Ile Ser Ala Phe Thr Leu Thr Asn Leu				
	415	420	425	
CTG CTT GTG GGT TTT GGC ATC ACC TGT CTC ATC AAC AAC TTA CAC CTC				1465
Leu Leu Val Gly Phe Gly Ile Thr Cys Leu Ile Asn Asn Leu His Leu				
	430	435	440	
CAG TTT GTG ACC TTT GTC CTG CAC ACC ATT GTT CGA GGT TTC TTC CAC				1513
Gln Phe Val Thr Phe Val Leu His Thr Ile Val Arg Gly Phe Phe His				
	445	450	455	460
TCA GCC TGT GGG AGT CTC TAT GCT GCA GTG TTC CCA TCC AAC CAC TTT				1561
Ser Ala Cys Gly Ser Leu Tyr Ala Ala Val Phe Pro Ser Asn His Phe				
	465	470	475	
GGG ACG CTG ACA GGC CTG CAG TCC CTC ATC AGT GCT GTG TTC GCC TTG				1609
Gly Thr Leu Thr Gly Leu Gln Ser Leu Ile Ser Ala Val Phe Ala Leu				
	480	485	490	
CTT CAG CAG CCA CTT TTC ATG GCG ATG GTG GGA CCC CTG AAA GGA GAG				1657
Leu Gln Gln Pro Leu Phe Met Ala Met Val Gly Pro Leu Lys Gly Glu				
	495	500	505	
CCC TTC TGG GTG AAT CTG GGC CTC CTG CTA TTC TCA CTC CTG GGA TTC				1705
Pro Phe Trp Val Asn Leu Gly Leu Leu Leu Phe Ser Leu Leu Gly Phe				
	510	515	520	
CTG TTG CCT TCC TAC CTC TTC TAT TAC CGT GCC CGG CTC CAG CAG GAG				1753
Leu Leu Pro Ser Tyr Leu Phe Tyr Tyr Arg Ala Arg Leu Gln Gln Glu				
	525	530	535	540
TAC GCC GCC AAT GGG ATG GGC CCA CTG AAG GTG CTT AGC GGC TCT GAG				1801
Tyr Ala Ala Asn Gly Met Gly Pro Leu Lys Val Leu Ser Gly Ser Glu				
	545	550	555	
GTG ACC GCA TAGACTTCTC AGACCAAGGG ACCTGGATGA				1840
Val Thr Ala				
CAGGCAATCA AGGCCTGAGC AACCAAAAGG AGTGCCCAT ATGGCTTTTC TACCTGTAAC				1900
ATGCACATAG AGCCATGGCC GTAGATTTAT AAATACCAAG AGAAGTTCTA TTTTGTAAA				1960
GACTGCAAAA AGGAGGAAAA AAAAACCTTC AAAAACGCCC CCTAAGTCAA CGCTCCATTG				2020
ACTGAAGACA GTCCCTATCC TAGAGGGGTT GAGCCTTCTT CCTCCTTGGG TTGGAGGAGA				2080
CCAGGGTGCC TCTTATCTCC TTCTAGCGGT CTGCCTCCTG GTACCTCTTG GGGGGATCGG				2140
CAACAGGCT ACCCCTGAGG TCCCATGTGC CATGAGTGTG CACACATGCA TGTGTCTGTG				2200
TATGTGTGAA TGTGAGAGAG ACACAGCCCT CTTTTCAGAA GGAAAGGGGC CTGAGGTGCC				2260

AGCTGTGTCC TGGGTTAGGG GTTGGGGGTC GGCCCCTTCC AGGGCCAGGA GGGCAGGTTC 2320  
 CCTCTCTGGT GCTGCTGCTT GCAAGTCTTA GAGGAAATAA AAAGGGAAGT GAG 2373

Sequence No.: 72

Sequence length: 1316

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Osteosarcoma

Cell line: U-2 OS

Clone name: HP10304

Sequence characteristics

Code representing characteristics: CDS

Existence site: 11.. 1003

Characterization method: E

Sequence description

GTTGTCCAAG ATG GAG GGC GCT CCA CCG GGG TCG CTC GCC CTC CGG CTC 49  
 Met Glu Gly Ala Pro Pro Gly Ser Leu Ala Leu Arg Leu  
 1 5 10  
 CTG CTG TTC GTG GCG CTA CCC GCC TCC GGC TGG CTG ACG ACG GGC GCC 97  
 Leu Leu Phe Val Ala Leu Pro Ala Ser Gly Trp Leu Thr Thr Gly Ala  
 15 20 25  
 CCC GAG CCG CCG CCG CTG TCC GGA GCC CCA CAG GAC GGC ATC AGA ATT 145  
 Pro Glu Pro Pro Pro Leu Ser Gly Ala Pro Gln Asp Gly Ile Arg Ile  
 30 35 40 45  
 AAT GTA ACT ACA CTG AAA GAT GAT GGG GAC ATA TCT AAA CAG CAG GTT 193  
 Asn Val Thr Thr Leu Lys Asp Asp Gly Asp Ile Ser Lys Gln Gln Val  
 50 55 60  
 GTT CTT AAC ATA ACC TAT GAG AGT GGA CAG GTG TAT GTA AAT GAC TTA 241  
 Val Leu Asn Ile Thr Tyr Glu Ser Gly Gln Val Tyr Val Asn Asp Leu  
 65 70 75  
 CCT GTA AAT AGT GGT GTA ACC CGA ATA AGC TGT CAG ACT TTG ATA GTG 289  
 Pro Val Asn Ser Gly Val Thr Arg Ile Ser Cys Gln Thr Leu Ile Val  
 80 85 90  
 AAG AAT GAA AAT CTT GAA AAT TTG GAG GAA AAA GAA TAT TTT GGA ATT 337  
 Lys Asn Glu Asn Leu Glu Asn Leu Glu Glu Lys Glu Tyr Phe Gly Ile  
 95 100 105  
 GTC AGT GTA AGG ATT TTA GTT CAT GAG TGG CCT ATG ACA TCT GGT TCC 385  
 Val Ser Val Arg Ile Leu Val His Glu Trp Pro Met Thr Ser Gly Ser

170

110	115	120	125	
AGT TTG CAA CTA ATT GTC ATT CAA GAA GAG GTA GTA GAG ATT GAT GGA				433
Ser Leu Gln Leu Ile Val Ile Gln Glu Glu Val Val Glu Ile Asp Gly				
	130	135	140	
AAA CAA GTT CAG CAA AAG GAT GTC ACT GAA ATT GAT ATT TTA GTT AAG				481
Lys Gln Val Gln Gln Lys Asp Val Thr Glu Ile Asp Ile Leu Val Lys				
	145	150	155	
AAC CGG GGA GTA CTC AGA CAT TCA AAC TAT ACC CTC CCT TTG GAA GAA				529
Asn Arg Gly Val Leu Arg His Ser Asn Tyr Thr Leu Pro Leu Glu Glu				
	160	165	170	
AGC ATG CTC TAC TCT ATT TCT CGA GAC AGT GAC ATT TTA TTT ACC CTT				577
Ser Met Leu Tyr Ser Ile Ser Arg Asp Ser Asp Ile Leu Phe Thr Leu				
	175	180	185	
CCT AAC CTC TCC AAA AAA GAA AGT GTT AGT TCA CTG CAA ACC ACT AGC				625
Pro Asn Leu Ser Lys Lys Glu Ser Val Ser Ser Leu Gln Thr Thr Ser				
	190	195	200	205
CAG TAT CTT ATC AGG AAT GTG GAA ACC ACT GTA GAT GAA GAT GTT TTA				673
Gln Tyr Leu Ile Arg Asn Val Glu Thr Thr Val Asp Glu Asp Val Leu				
	210	215	220	
CCT GGC AAG TTA CCT GAA ACT CCT CTC AGA GCA GAG CCG CCA TCT TCA				721
Pro Gly Lys Leu Pro Glu Thr Pro Leu Arg Ala Glu Pro Pro Ser Ser				
	225	230	235	
TAT AAG GTA ATG TGT CAG TGG ATG GAA AAG TTT AGA AAA GAT CTG TGT				769
Tyr Lys Val Met Cys Gln Trp Met Glu Lys Phe Arg Lys Asp Leu Cys				
	240	245	250	
AGG TTC TGG AGC AAC GTT TTC CCA GTA TTC TTT CAG TTT TTG AAC ATC				817
Arg Phe Trp Ser Asn Val Phe Pro Val Phe Phe Gln Phe Leu Asn Ile				
	255	260	265	
ATG GTG GTT GGA ATT ACA GGA GCA GCT GTG GTA ATA ACC ATC TTA AAG				865
Met Val Val Gly Ile Thr Gly Ala Ala Val Val Ile Thr Ile Leu Lys				
	270	275	280	285
GTG TTT TTC CCA GTT TCT GAA TAC AAA GGA ATT CTT CAG TTG GAT AAA				913
Val Phe Phe Pro Val Ser Glu Tyr Lys Gly Ile Leu Gln Leu Asp Lys				
	290	295	300	
GTG GAC GTC ATA CCT GTG ACA GCT ATC AAC TTA TAT CCA GAT GGT CCA				961
Val Asp Val Ile Pro Val Thr Ala Ile Asn Leu Tyr Pro Asp Gly Pro				
	305	310	315	
GAG AAA AGA GCT GAA AAC CTT GAA GAT AAA ACA TGT ATT TAAAACGCCA				1010
Glu Lys Arg Ala Glu Asn Leu Glu Asp Lys Thr Cys Ile				
	320	325	330	
TCTCATATCA TGGACTCCGA AGTAGCCTGT TGCCTCCAAA TTTGCCACTT GAATATAATT				1070
TTCTTTAAAT CGTTAAGAAT CAGTTTATAC ACTAGAGAAA TTGCTAAACT CTAAGACTGC				1130
CTGAAAATTG ACCTTTACAG TGCCAAGTTA AAGTTTACCT TATTCTCGGC CGGGTGCAGT				1190
GGCTCATGCC TGTAATCCCA GGACTTTGGG AGGCCAATGC GGGCGGATCA CGAGGTCAGA				1250



TCAAGACCAT CCTGCCAACA TGGTGAAACC CTGTCTCTAC TAAAAAAAAT AAAAAAGTTA 1310  
GCTGGG 1316

Sequence No.: 73

Sequence length: 893

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Osteosarcoma

Cell line: U-2 OS

Clone name: HP10305

Sequence characteristics

Code representing characteristics: CDS

Existence site: 110.. 436

Characterization method: E

Sequence description

ATCGCGGAGT CGGTGCTTTA GTACGCCGCT GGCACCTTTA CTCTGCCCGG CCGCGCGAAC 60  
CCGTTTGAGC TCGGTATCCT AGTGCACACG CCTTGCAAGC GACGGCGCC ATG AGT CTG 118  
Met Ser Leu  
1  
ACT TCC AGT TCC AGC GTA CGA GTT GAA TGG ATC GCA GCA GTT ACC ATT 166  
Thr Ser Ser Ser Ser Val Arg Val Glu Trp Ile Ala Ala Val Thr Ile  
5 10 15  
GCT GCT GGG ACA GCT GCA ATT GGT TAT CTA GCT TAC AAA AGA TTT TAT 214  
Ala Ala Gly Thr Ala Ala Ile Gly Tyr Leu Ala Tyr Lys Arg Phe Tyr  
20 25 30 35  
GTT AAA GAT CAT CGA AAT AAA GCT ATG ATA AAC CTT CAC ATC CAG AAA 262  
Val Lys Asp His Arg Asn Lys Ala Met Ile Asn Leu His Ile Gln Lys  
40 45 50  
GAC AAC CCC AAG ATA GTA CAT GCT TTT GAC ATG GAG GAT TTG GGA GAT 310  
Asp Asn Pro Lys Ile Val His Ala Phe Asp Met Glu Asp Leu Gly Asp  
55 60 65  
AAA GCT GTG TAC TGC CGT TGT TGG AGG TCC AAA AAG TTC CCA TTC TGT 358  
Lys Ala Val Tyr Cys Arg Cys Trp Arg Ser Lys Lys Phe Pro Phe Cys  
70 75 80  
GAT GGG GCT CAC ACA AAA CAT AAC GAA GAG ACT GGA GAC AAT GTG GGC 406  
Asp Gly Ala His Thr Lys His Asn Glu Glu Thr Gly Asp Asn Val Gly  
85 90 95  
CCT CTG ATC ATC AAG AAA AAA GAA ACT TAAATGGACA CTTTGA 450

Pro Leu Ile Ile Lys Lys Lys Glu Thr

100	105		
TGCTGCAAAAT CAGCTTGTCG TGAAGTTACC TGATTGTTTA ATTAGAATGA CTACCACCTC	510		
TGTCTGATTC ACCTTCGCTG GATTCTAAAT GTGGTATATT GCAAACCTGCA GCTTTCACAT	570		
TTATGGCATT TGTCTTGTTG AAACATCGTG GTGCACATTT GTTTAAACAA AAAAAAAAAA	630		
AAAAAGGAAA AACCAACCTC ATGGCCTGTG GGTTATTTTG GTCTTGTAAG GATCCATTTT	690		
TTTAAATAC TGACATATAG AGTTGTACCT TATATAGAAT ATAGTTGTAT CTTGAAGTCA	750		
ACATATTAAA TTATTCTCAA AATTATGTAT TTGCAGATTG TACTTGTAAG TTTCAAAGAA	810		
AAATTACCAT CTTTTCATAT TGACCTGGAA ACTAAATAGG ATGTGATTCA GCTACATTAA	870		
TTTCTTAATA CAATCTAGGA AAG	893		

Sequence No.: 74

Sequence length: 690

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Osteosarcoma

Cell line: U-2 OS

Clone name: HP10306

Sequence characteristics

Code representing characteristics: CDS

Existence site: 230.. 535

Characterization method: E

Sequence description

TAACAGCGCA TCGTGCACT GTTGCTCGC CCAAAGAAGA CTACAATCTC CAGGGAAACC	60
TGGGGCGTCT CGCGCAAACG TCCATAACTG AAAGTAGCTA AGGCACCCCA GCCGGAGGAA	120
GTGAGCTCTC CTGGGGCGTG GTTGTTCTGT ATCCTTGCAT CTGTTACTTA GGGTCAAGGC	180
TTGGGTCTTG CCCCGCAGAC CCTTGGGACG ACCCGGCCCC AGCGCAGCT ATG AAC CTG	238
Met Asn Leu	
1	
GAG CGA GTG TCC AAT GAG GAG AAA TTG AAC CTG TGC CGG AAG TAC TAC	286
Glu Arg Val Ser Asn Glu Glu Lys Leu Asn Leu Cys Arg Lys Tyr Tyr	
5 10 15	
CTG GGG GGG TTT GCT TTC CTG CCT TTT CTC TGG TTG GTC AAC ATC TTC	334
Leu Gly Gly Phe Ala Phe Leu Pro Phe Leu Trp Leu Val Asn Ile Phe	
20 25 30 35	
TGG TTC TTC CGA GAG GCC TTC CTT GTC CCA GCC TAC ACA GAA CAG AGC	382
Trp Phe Phe Arg Glu Ala Phe Leu Val Pro Ala Tyr Thr Glu Gln Ser	
40 45 50	

173

CAA ATC AAA GGC TAT GTC TGG CGC TCA GCT GTG GGC TTC CTC TTC TGG	430
Gln Ile Lys Gly Tyr Val Trp Arg Ser Ala Val Gly Phe Leu Phe Trp	
55 60 65	
GTG ATA GTG CTC ACC TCC TGG ATC ACC ATC TTC CAG ATC TAC CGG CCC	478
Val Ile Val Leu Thr Ser Trp Ile Thr Ile Phe Gln Ile Tyr Arg Pro	
70 75 80	
CGC TGG GGT GCC CTT GGG GAC TAC CTC TCC TTC ACC ATA CCC CTG GGC	526
Arg Trp Gly Ala Leu Gly Asp Tyr Leu Ser Phe Thr Ile Pro Leu Gly	
85 90 95	
ACC CCC TGACAACTTC TGCACATACT GGGGCCCTGC TTATTCTCCC AGGACAGG	580
Thr Pro	
100	
CTCCTTAAAG CAGAGGAGCC TGTCCTGGGA GCCCCTTCTC AAACCTCCTAA GACTTGTTTT	640
CATGTCCAC GTTCTCTGCT GACATCCCCC AATAAAGGAC CCTAACTTTC	690

Sequence No.: 75

Sequence length: 2186

Sequence type: Nucleic acid

Strandedness: Double

Topology: Linear

Sequence kind: cDNA to mRNA

Original source:

Organism species: *Homo sapiens*

Cell kind: Epidermoid carcinoma

Cell line: KB

Clone name: HP10328

Sequence characteristics

Code representing characteristics: CDS

Existence site: 118.. 1236

Characterization method: E

Sequence description

ACTCTTTCTT CGGCTCGCGA GCTGAGAGGA GCAGGTAGAG GGGCAGAGGC GGGACTGTCC	60
TCTGGGGGAG CCGCCAGGA GGCTCCTCAG GCCGACCCCA GACCCTGGCT GGCCAGG	117
ATG AAG TAT CTC CGG CAC CGG CGG CCC AAT GCC ACC CTC ATT CTG GCC	165
Met Lys Tyr Leu Arg His Arg Arg Pro Asn Ala Thr Leu Ile Leu Ala	
1 5 10 15	
ATC GGC GCT TTC ACC CTC CTC CTC TTC AGT CTG CTA GTG TCA CCA CCC	213
Ile Gly Ala Phe Thr Leu Leu Leu Phe Ser Leu Leu Val Ser Pro Pro	
20 25 30	
ACC TGC AAG GTC CAG GAG CAG CCA CCG GCG ATC CCC GAG GCC CTG GCC	261
Thr Cys Lys Val Gln Glu Gln Pro Pro Ala Ile Pro Glu Ala Leu Ala	
35 40 45	

TGG CCC ACT CCA CCC ACC CGC CCA GCC CCG GCC CCG TGC CAT GCC AAC	309
Trp Pro Thr Pro Pro Thr Arg Pro Ala Pro Ala Pro Cys His Ala Asn	
50 55 60	
ACC TCT ATG GTC ACC CAC CCG GAC TTC GCC ACG CAG CCG CAG CAC GTT	357
Thr Ser Met Val Thr His Pro Asp Phe Ala Thr Gln Pro Gln His Val	
65 70 75 80	
CAG AAC TTC CTC CTG TAC AGA CAC TGC CGC CAC TTT CCC CTG CTG CAG	405
Gln Asn Phe Leu Leu Tyr Arg His Cys Arg His Phe Pro Leu Leu Gln	
85 90 95	
GAC GTG CCC CCC TCT AAG TGC GCG CAG CCG GTC TTC CTG CTG CTG GTG	453
Asp Val Pro Pro Ser Lys Cys Ala Gln Pro Val Phe Leu Leu Leu Val	
100 105 110	
ATC AAG TCC TCC CCT AGC AAC TAT GTG CGC CGC GAG CTG CTG CGG CGC	501
Ile Lys Ser Ser Pro Ser Asn Tyr Val Arg Arg Glu Leu Leu Arg Arg	
115 120 125	
ACG TGG GGC CGC GAG CGC AAG GTA CGG GGT TTG CAG CTG CGC CTC CTC	549
Thr Trp Gly Arg Glu Arg Lys Val Arg Gly Leu Gln Leu Arg Leu Leu	
130 135 140	
TTC CTG GTG GGC ACA GCC TCC AAC CCG CAC GAG GCC CGC AAG GTC AAC	597
Phe Leu Val Gly Thr Ala Ser Asn Pro His Glu Ala Arg Lys Val Asn	
145 150 155 160	
CGG CTG CTG GAG CTG GAG GCA CAG ACT CAC GGA GAC ATC CTG CAG TGG	645
Arg Leu Leu Glu Leu Glu Ala Gln Thr His Gly Asp Ile Leu Gln Trp	
165 170 175	
GAC TTC CAC GAC TCC TTC TTC AAC CTC ACG CTC AAG CAG GTC CTG TTC	693
Asp Phe His Asp Ser Phe Phe Asn Leu Thr Leu Lys Gln Val Leu Phe	
180 185 190	
TTA CAG TGG CAG GAG ACA AGG TGC GCC AAC GCC AGC TTC GTG CTC AAC	741
Leu Gln Trp Gln Glu Thr Arg Cys Ala Asn Ala Ser Phe Val Leu Asn	
195 200 205	
GGG GAT GAT GAC GTC TTT GCA CAC ACA GAC AAC ATG GTC TTC TAC CTG	789
Gly Asp Asp Asp Val Phe Ala His Thr Asp Asn Met Val Phe Tyr Leu	
210 215 220	
CAG GAC CAT GAC CCT GGC CGC CAC CTC TTC GTG GGG CAA CTG ATC CAA	837
Gln Asp His Asp Pro Gly Arg His Leu Phe Val Gly Gln Leu Ile Gln	
225 230 235 240	
AAC GTG GGC CCC ATC CGG GCT TTT TGG AGC AAG TAC TAT GTG CCA GAG	885
Asn Val Gly Pro Ile Arg Ala Phe Trp Ser Lys Tyr Tyr Val Pro Glu	
245 250 255	
GTG GTG ACT CAG AAT GAG CGG TAC CCA CCC TAT TGT GGG GGT GGT GGC	933
Val Val Thr Gln Asn Glu Arg Tyr Pro Pro Tyr Cys Gly Gly Gly Gly	
260 265 270	
TTC TTG CTG TCC CGC TTC ACG GCC GCT GCC CTG CGC CGT GCT GCC CAT	981
Phe Leu Leu Ser Arg Phe Thr Ala Ala Ala Leu Arg Arg Ala Ala His	

275	280	285	
GTC TTG GAC ATC TTC CCC ATT GAT GAT GTC TTC CTG GGT ATG TGT CTG			1029
Val Leu Asp Ile Phe Pro Ile Asp Asp Val Phe Leu Gly Met Cys Leu			
290	295	300	
GAG CTT GAG GGA CTG AAG CCT GCC TCC CAC AGC GGC ATC CGC ACG TCT			1077
Glu Leu Glu Gly Leu Lys Pro Ala Ser His Ser Gly Ile Arg Thr Ser			
305	310	315	320
GGC GTG CGG GCT CCA TCG CAA CAC CTG TCC TCC TTT GAC CCC TGC TTC			1125
Gly Val Arg Ala Pro Ser Gln His Leu Ser Ser Phe Asp Pro Cys Phe			
325	330	335	
TAC CGA GAC CTG CTG CTG GTG CAC CGC TTC CTA CCT TAT GAG ATG CTG			1173
Tyr Arg Asp Leu Leu Leu Val His Arg Phe Leu Pro Tyr Glu Met Leu			
340	345	350	
CTC ATG TGG GAT GCG CTG AAC CAG CCC AAC CTC ACC TGC GGC AAT CAG			1221
Leu Met Trp Asp Ala Leu Asn Gln Pro Asn Leu Thr Cys Gly Asn Gln			
355	360	365	
ACA CAG ATC TAC TGAGTCAGCA TCAGGGTCCC CAGCCTCTGG GCTCCTG			1270
Thr Gln Ile Tyr			
370			
TTTCCATAGG AAGGGGCGAC ACCTTCCTCC CAGGAAGCTG AGACCTTTGT GGTCTGAGCA			1330
TAAGGGAGTG CCAGGGAAGG TTTGAGGTTT GATGAGTGAA TATTCTGGCT GGCGAACTCC			1390
TACACATCCT TCAAAACCCA CCTGGTACTG TTCCAGCATC TTCCCTGGAT GGCTGGAGGA			1450
ACTCCAGAAA ATATCCATCT TCTTTTTGTG GCTGCTAATG GCAGAAAGTG CTGTGCTAGA			1510
GTTCCAAC TG GATGCATC CGTCCCGTTT GAGTCAAAGT CTTACTTCCC TGCTCTCACC			1570
TACTCACAGA CGGGATGCTA AGCAGTGCAC CTGCAGTGGT TTAATGGCAG ATAAGCTCCG			1630
TCTGCAGTTC CAGGCCAGCC AGAAACTCCT GTGTCCACAT AGAGCTGACG TGAGAAATAT			1690
CTTTCAGCCC AGCAGAGAGG GGTCTGATC TTAACCCTTT CCTGGGTCTC AGACAACTCA			1750
GAAGGTTGGG GGGATACCAG AGAGGTGGTG GAATAGGACC GCCCCCTCCT TACTTGTGGG			1810
ATCAAATGCT GTAATGGTGG AGGTGTGGGC AGAGGAGGGA GGCAAGTGTC CTTTGAAAGT			1870
TGTGAGAGCT CAGAGTTTCT GGGGTCTCTA TTAGGAGCCC CCATCCCTGT GTTCCCCAAG			1930
AATTCAGAGA ACAGCACTGG GGCTGGAATG ATCTTTAATG GGCCCAAGGC CAACAGGCAT			1990
ATGCCTCACT ACTGCCTGGA GAAGGGAGAG ATTCAAGTCC TCCAGCAGCC TCCCTCACCC			2050
AGTATGTTTT ACAGATTACG GGGGGACCGG GTGAGCCAGT GACCCCTGCG AGCCCCCAGC			2110
TTCAGGCCTC AGTGTCTGCC AGTCAAGCTT CACAGGCATT GTGATGGGGC AGCCTTGGGG			2170
AATATAAAAT TTTGTG			2186

## Claims

1. A protein containing any of the amino acid sequences represented by Sequence No. 1 to Sequence No. 2 or by Sequence No. 4 to Sequence No. 25.
2. A DNA encoding any of the proteins as described in Claim 1.
3. A cDNA containing any of the base sequences represented by Sequence No. 26 to Sequence No. 50.
4. A cDNA as described in Claim 3 which comprises any of the base sequences represented by Sequence No. 51 to Sequence No. 75.
5. A transformed eukaryotic cell capable of expressing any of DNAs as described in Claim 2 to 4 and producing a protein as described in Claim 1.

1/28

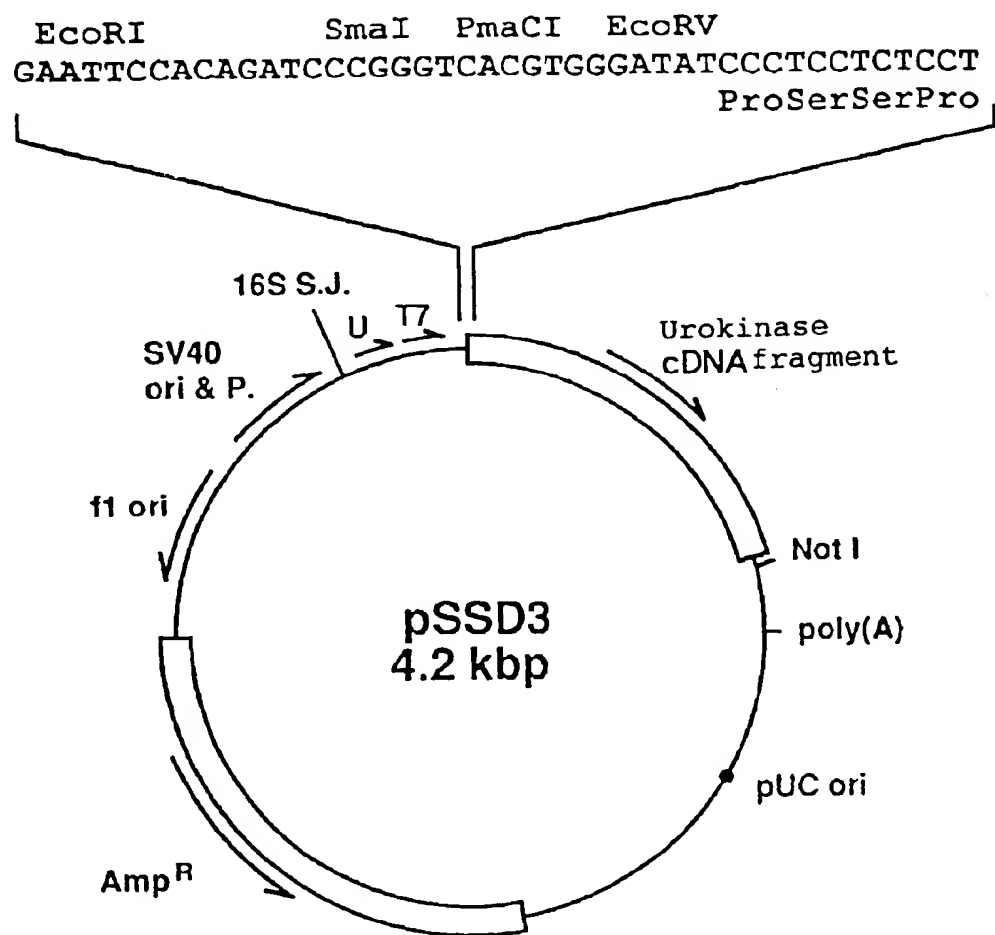
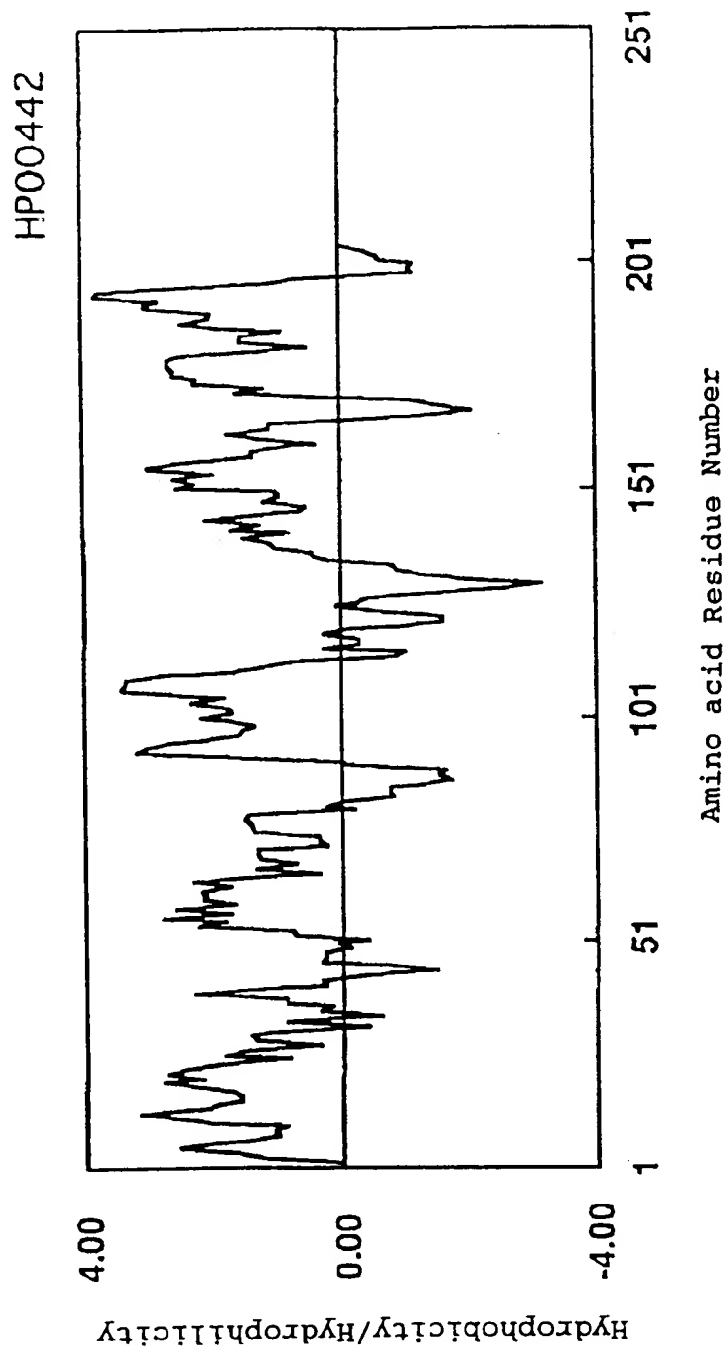


Fig. 1

2/28

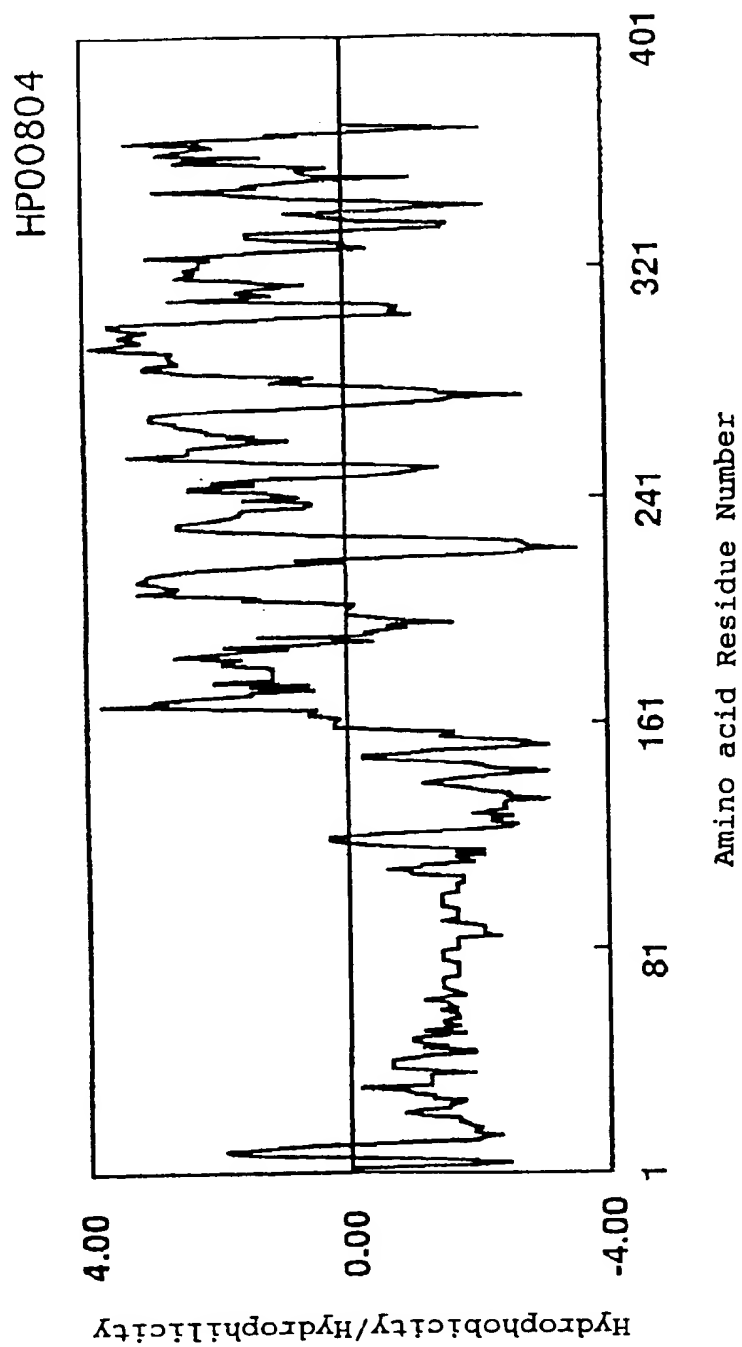
Fig. 2





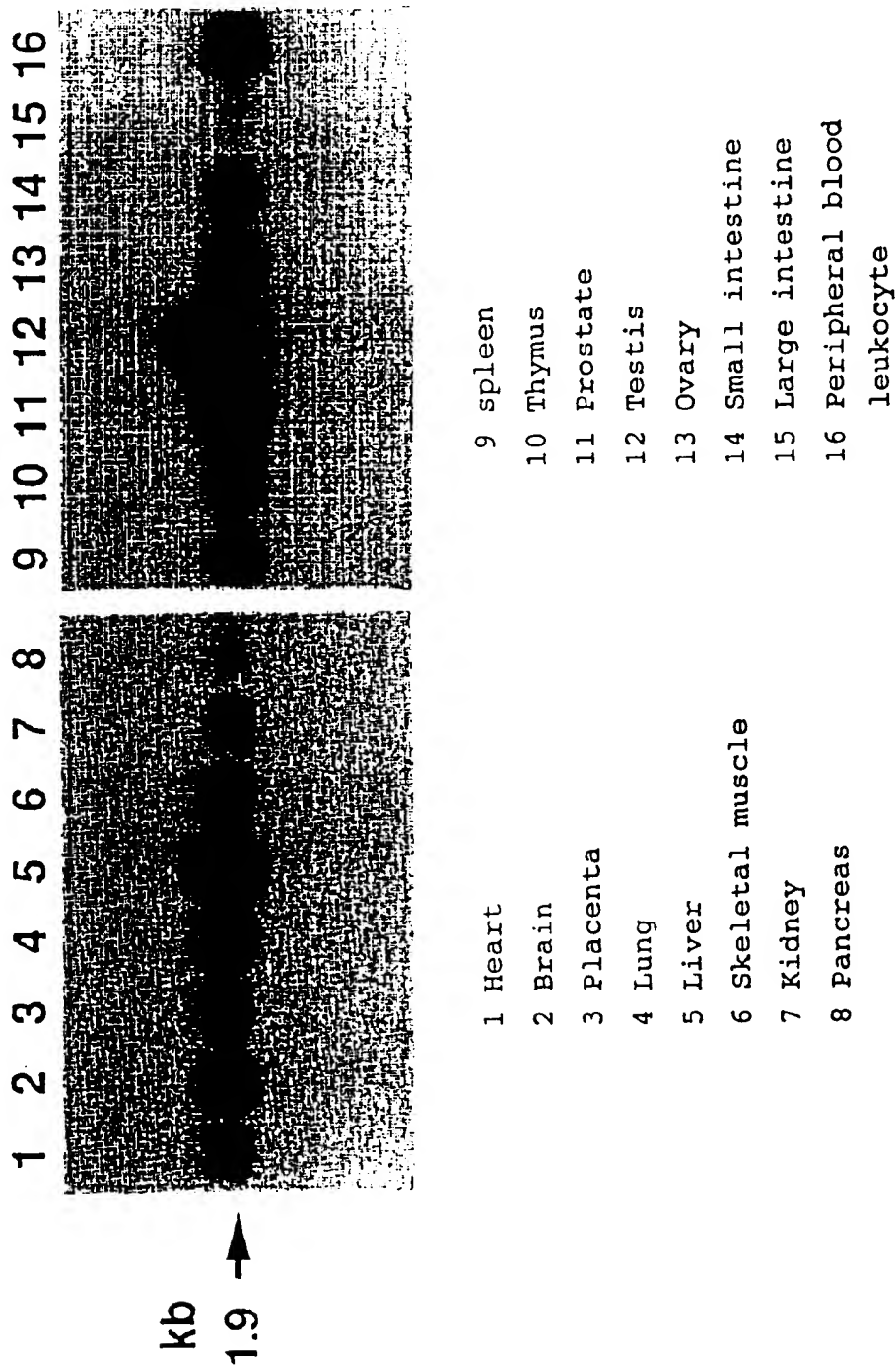
3/28

Fig. 3



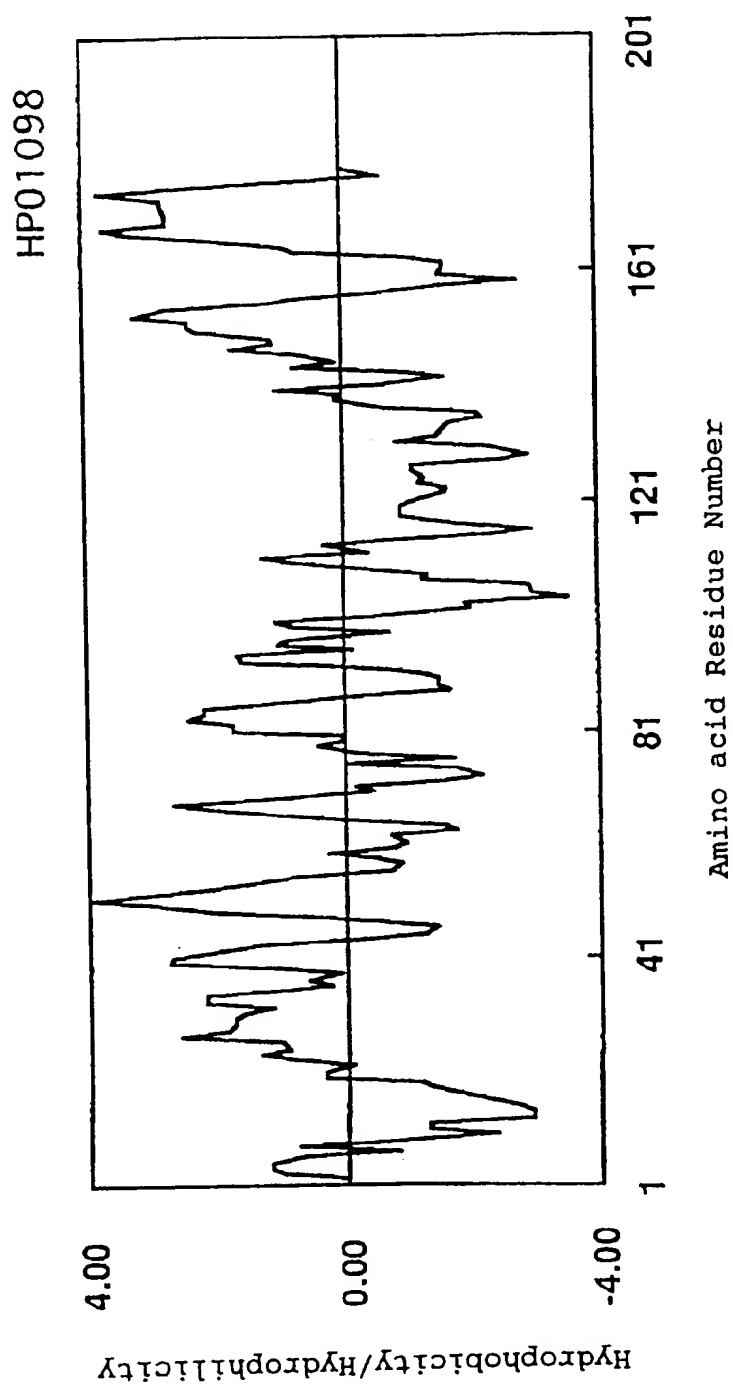
4/28

Fig. 4



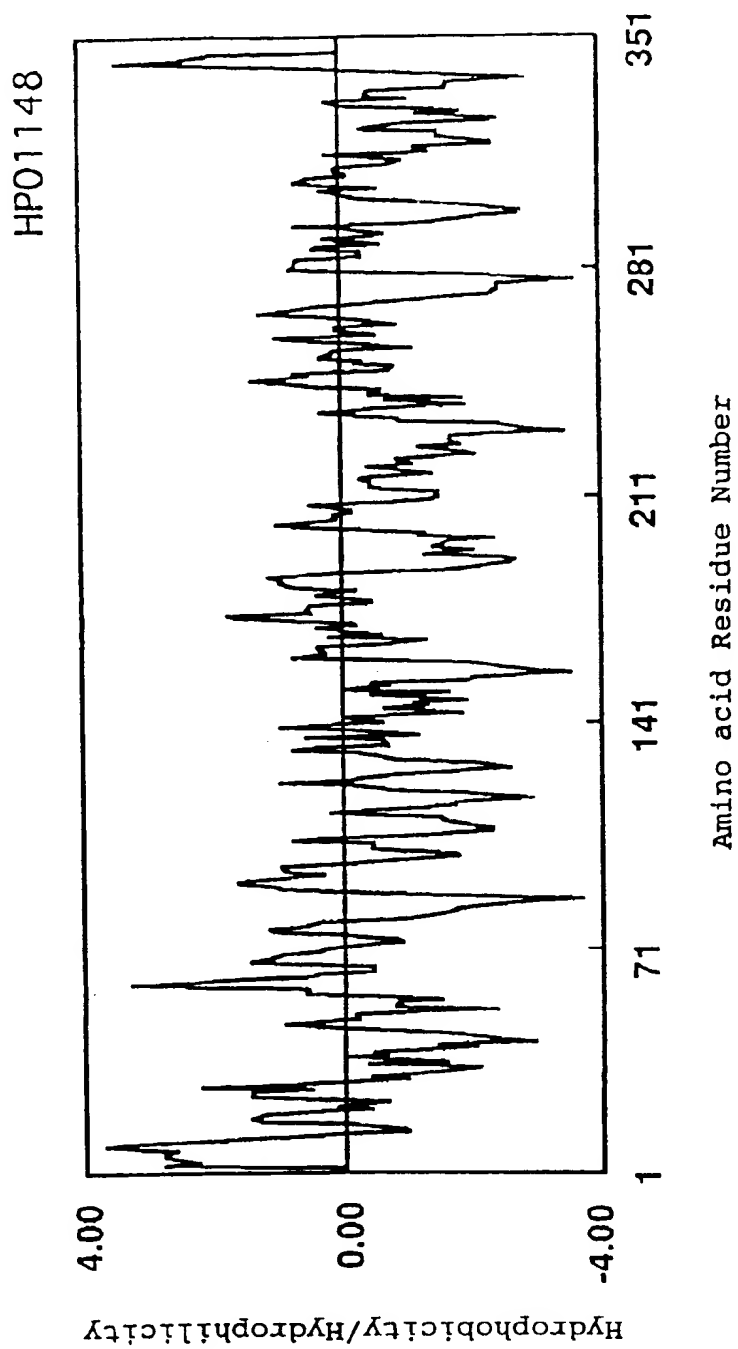
5/28

Fig. 5



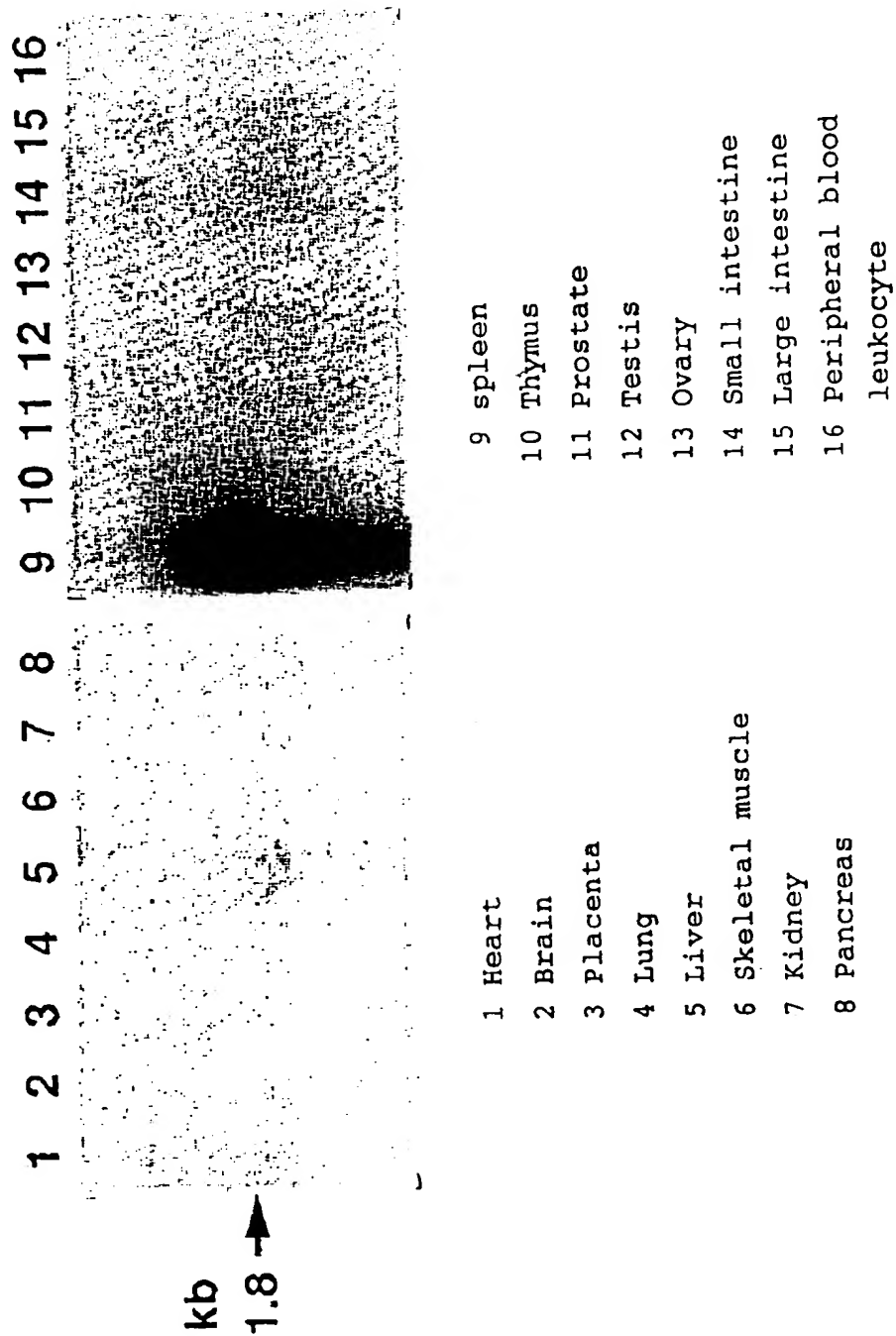
6/28

Fig. 6



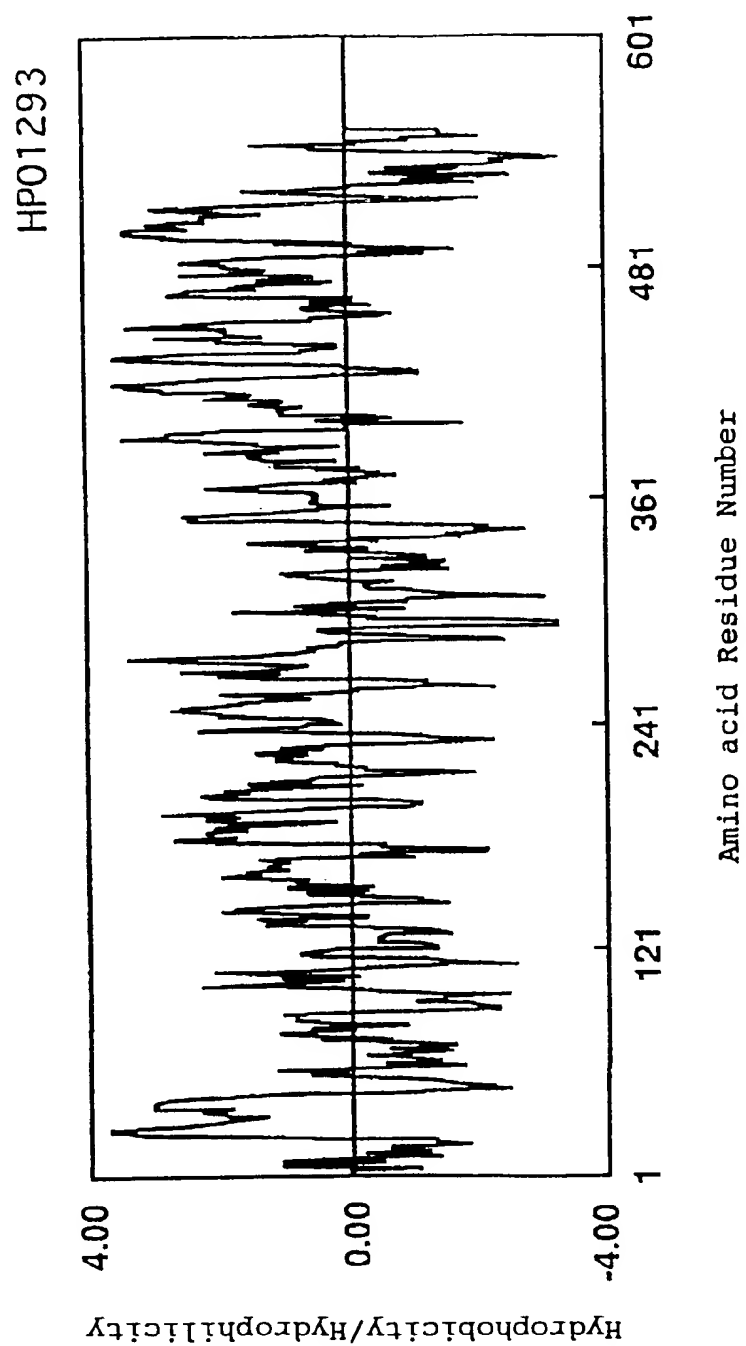
7/28

Fig. 7



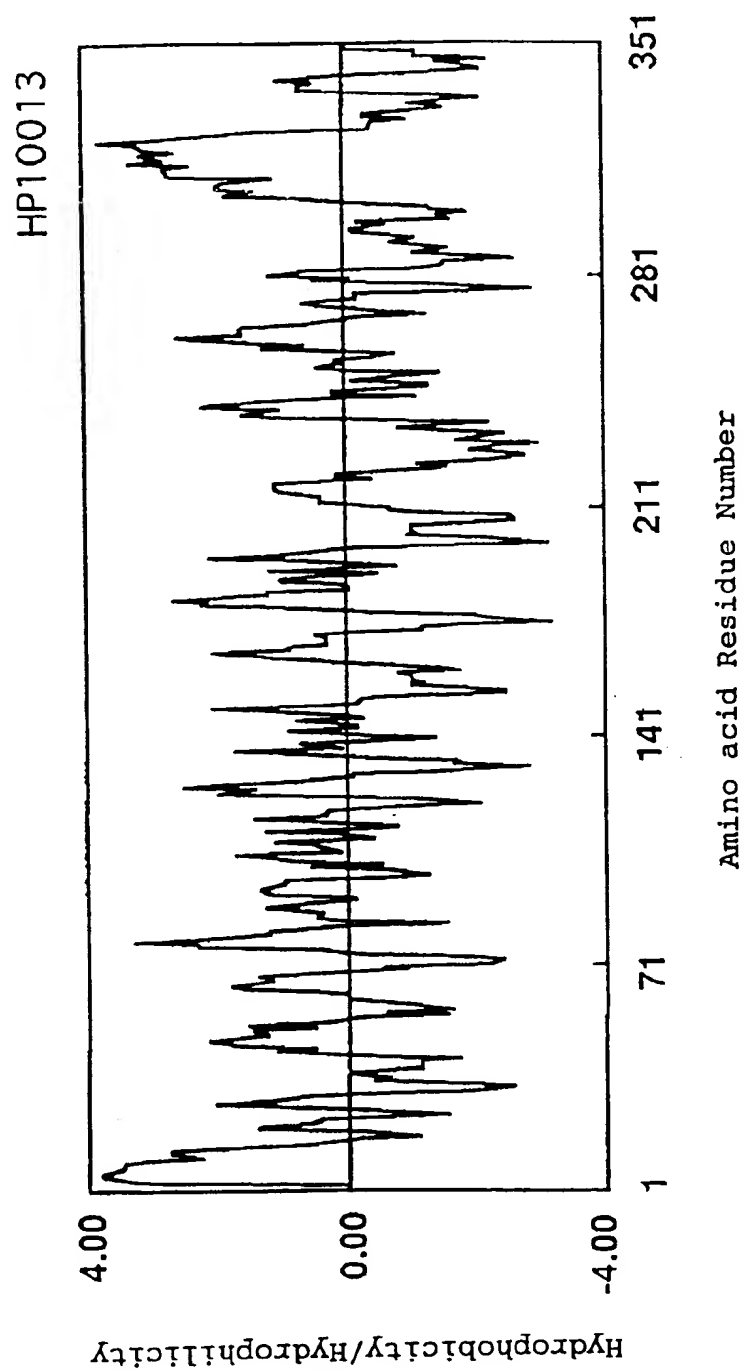
8/28

Fig. 8



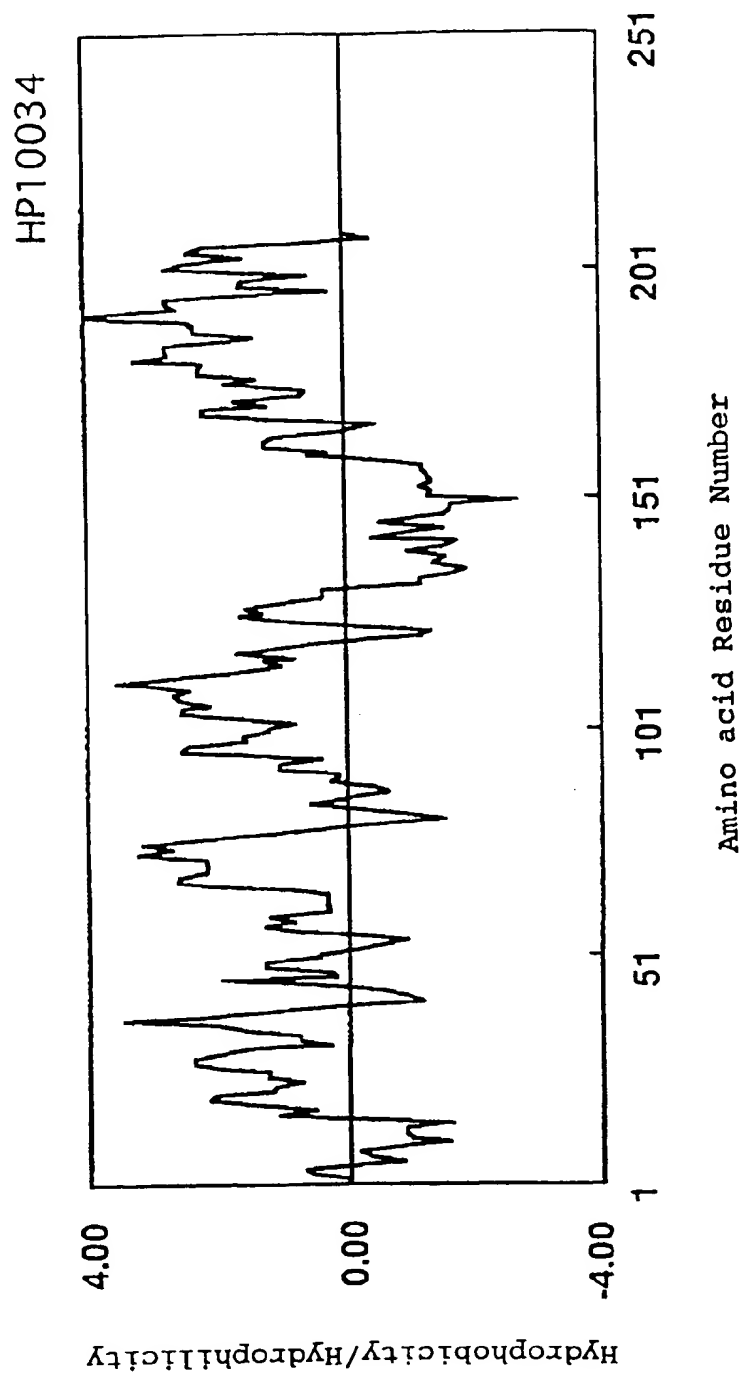
9/28

Fig. 9



10/28

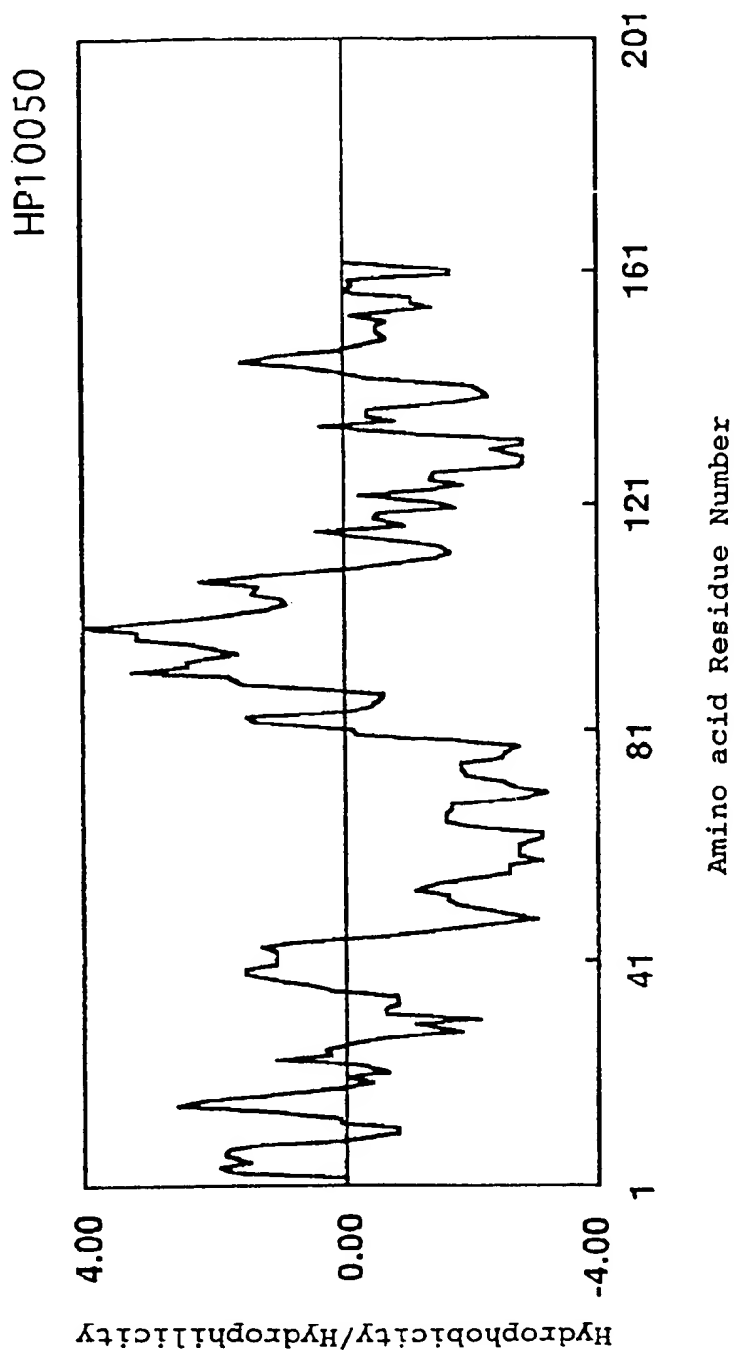
Fig. 10





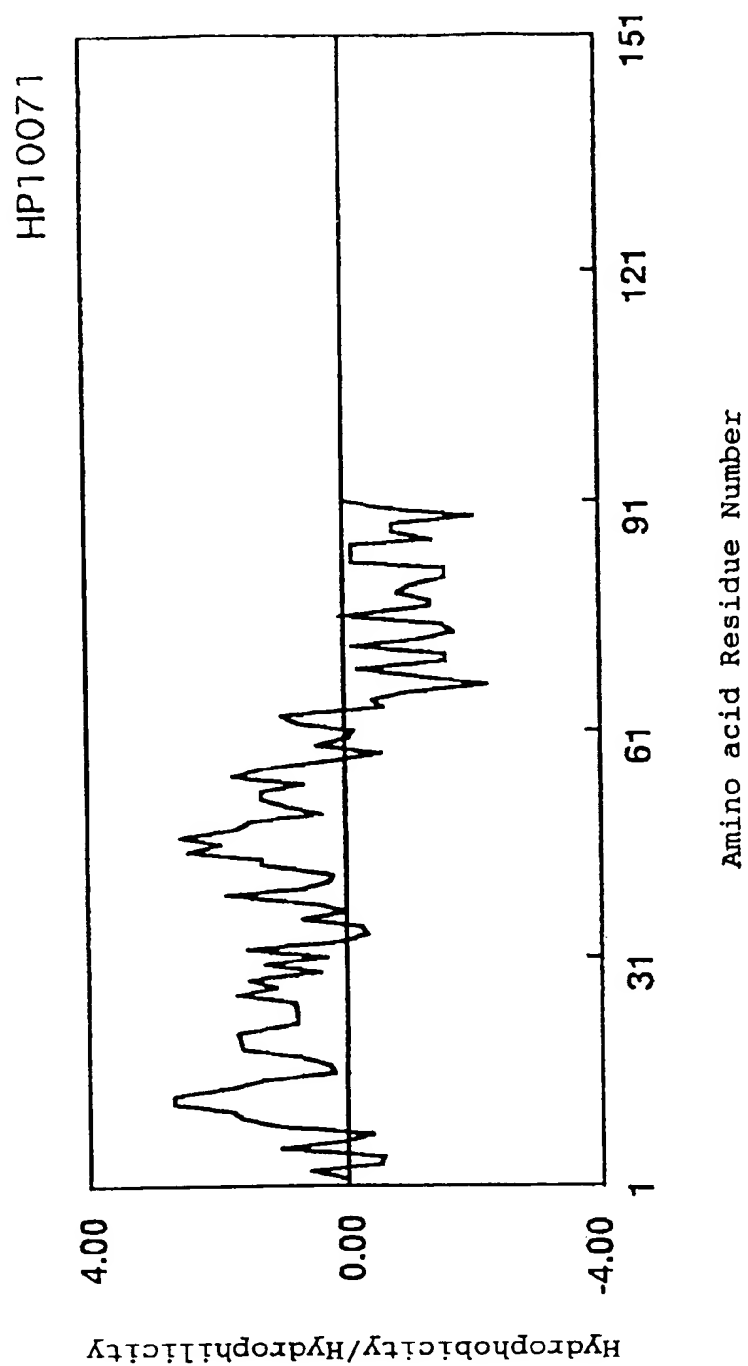
11/28

Fig. 11



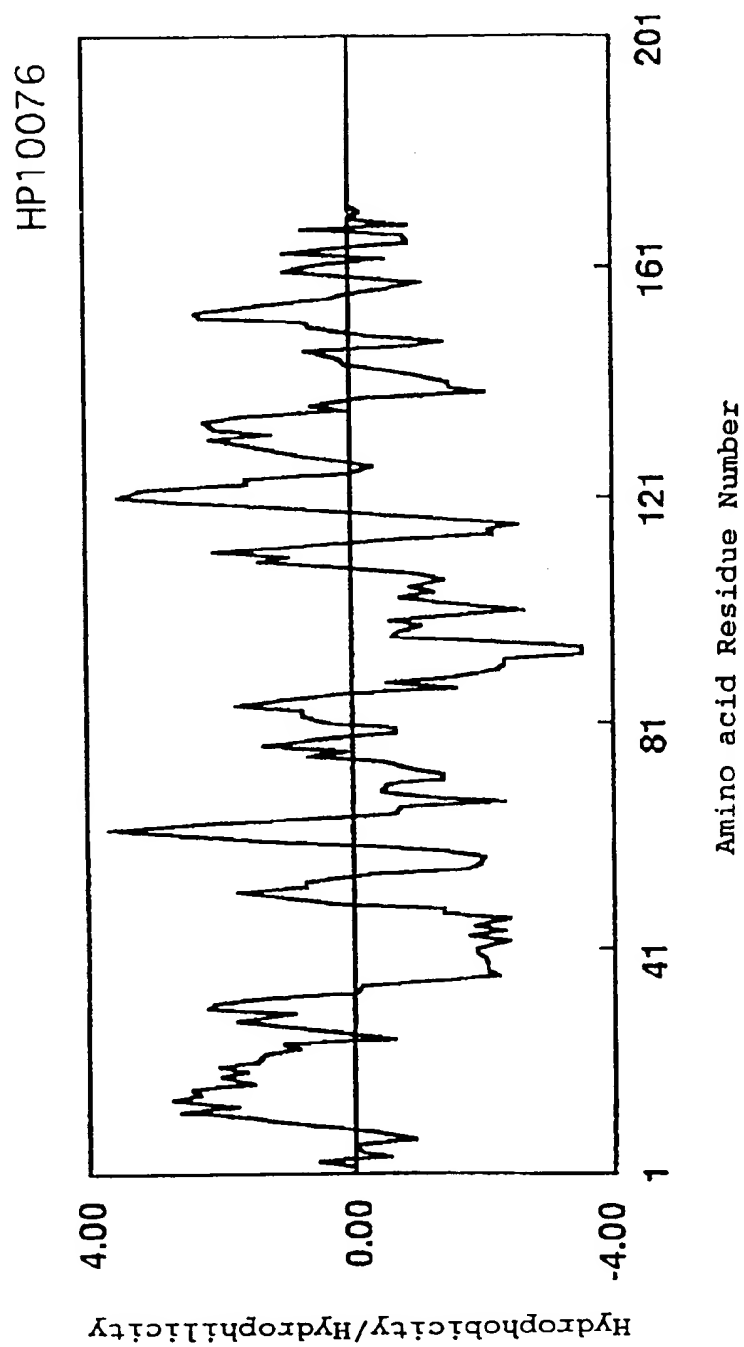
12/28

Fig. 12



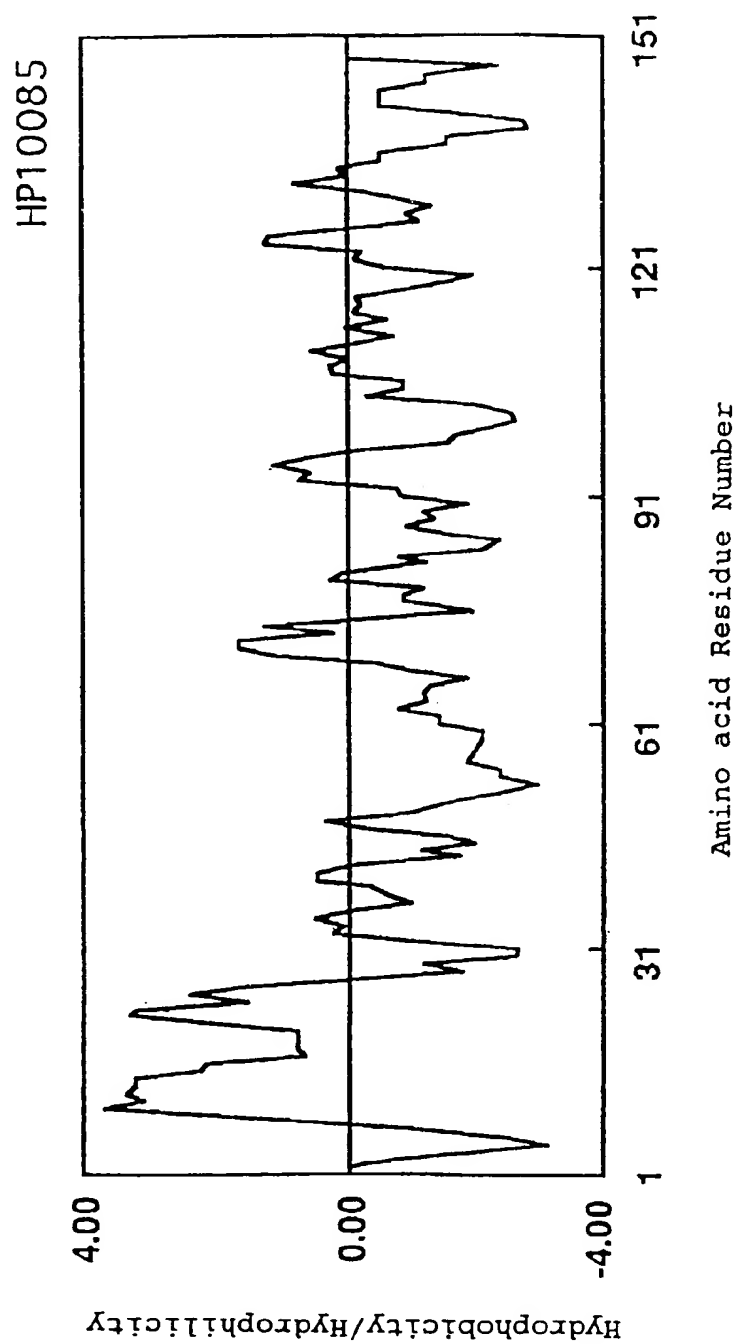
13/28

Fig. 13



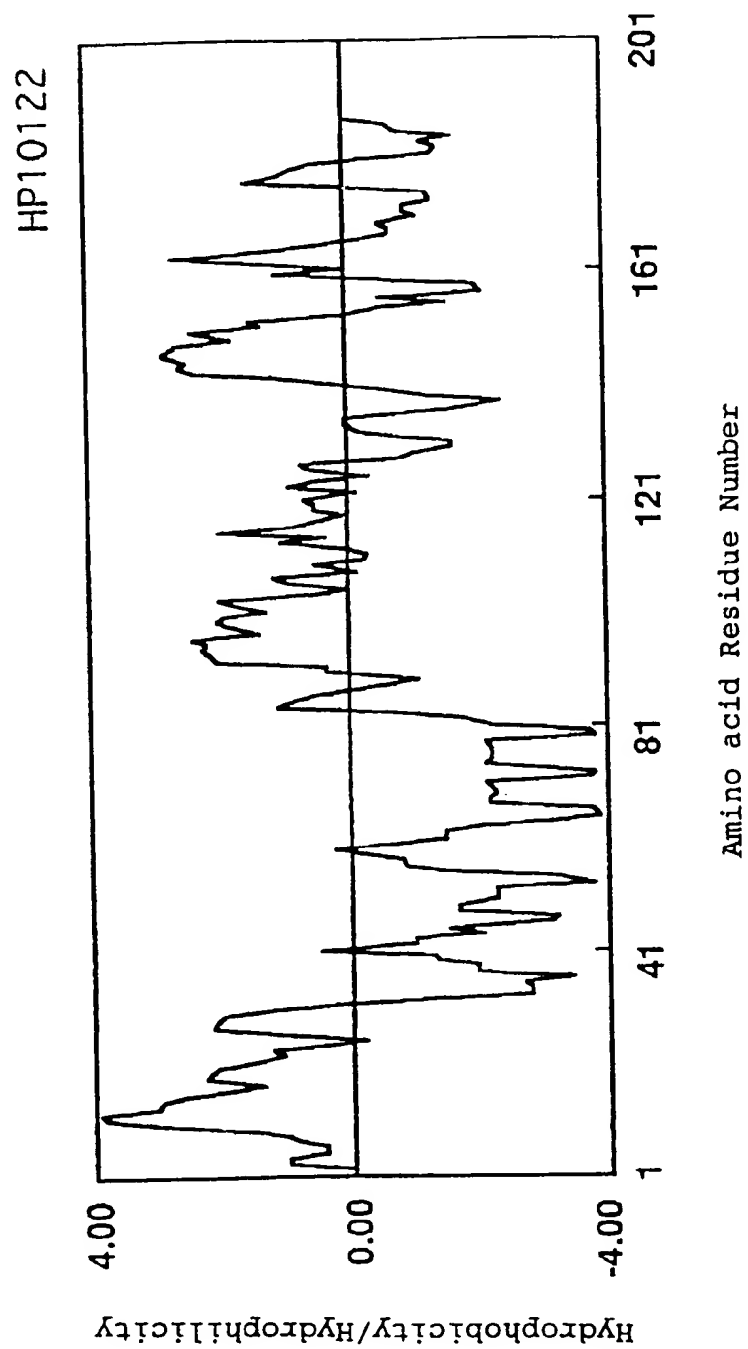
14/28

Fig. 14



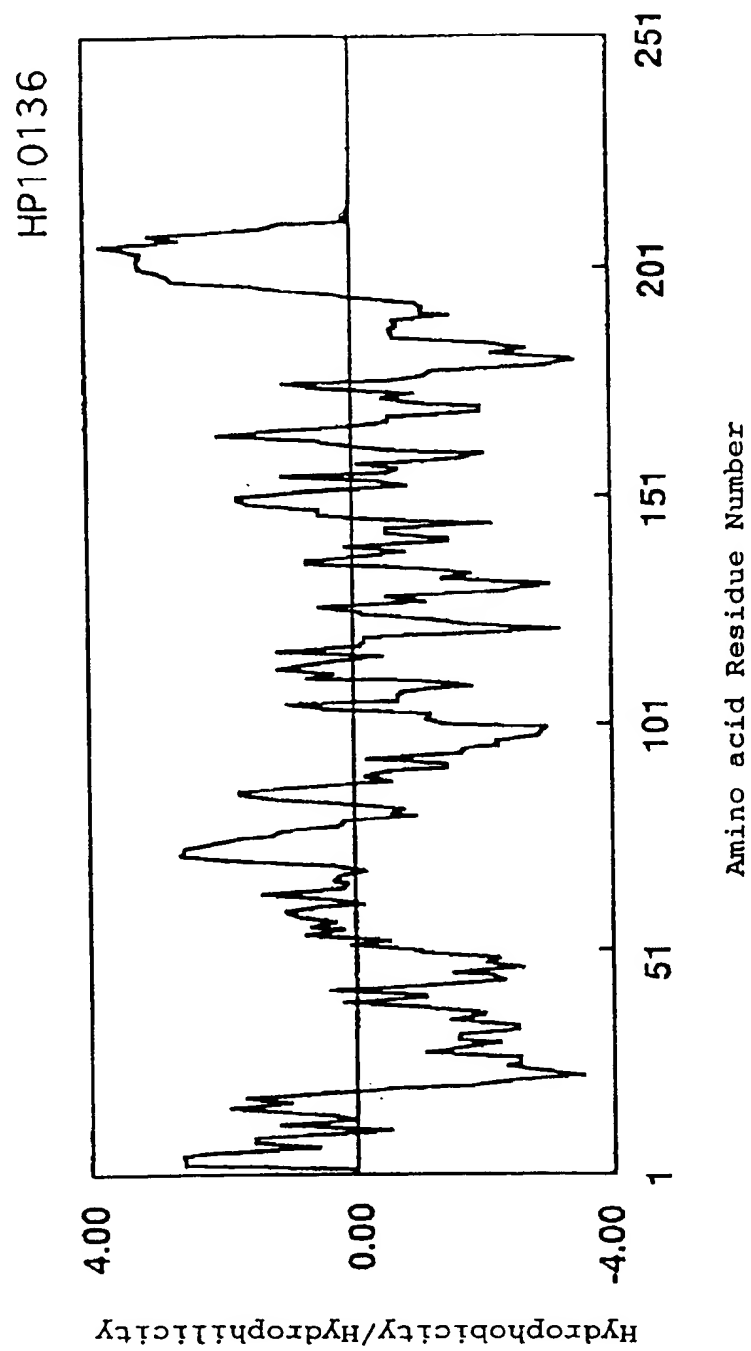
15/28

Fig. 15



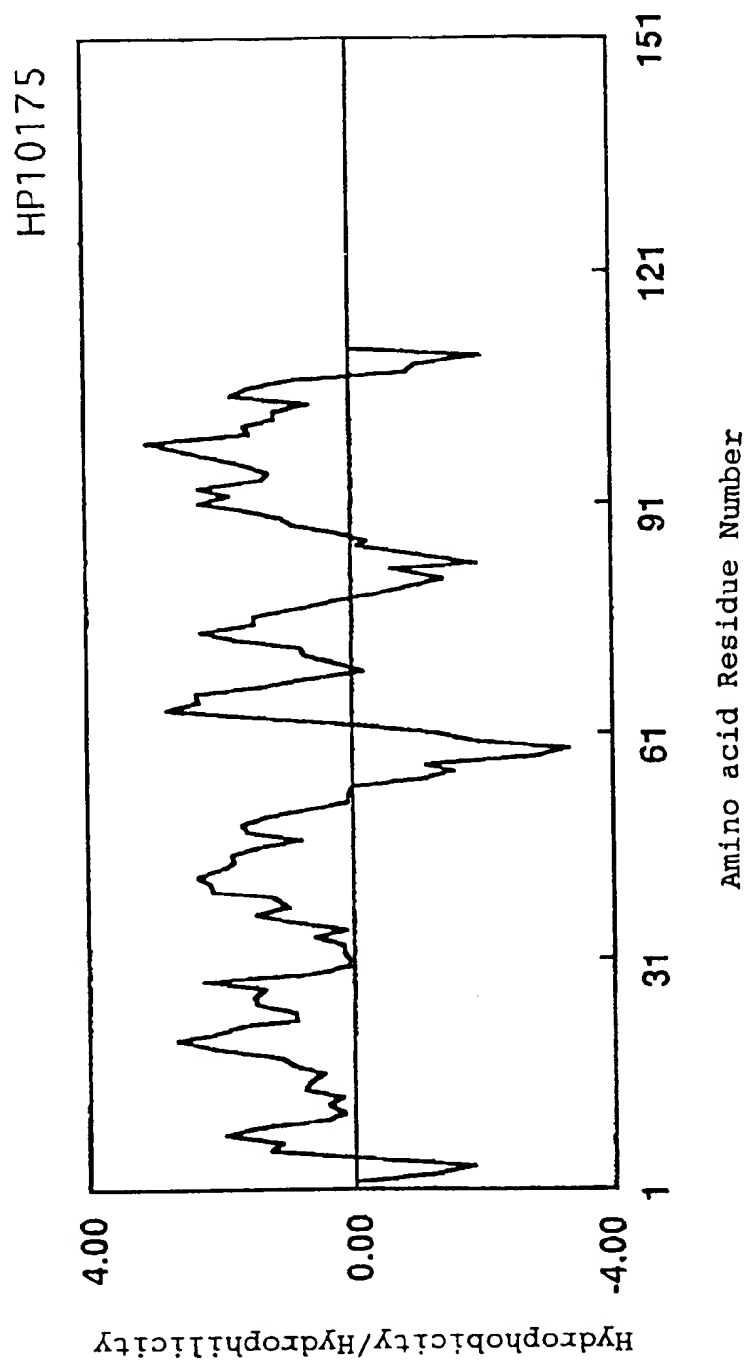
16/28

Fig. 16



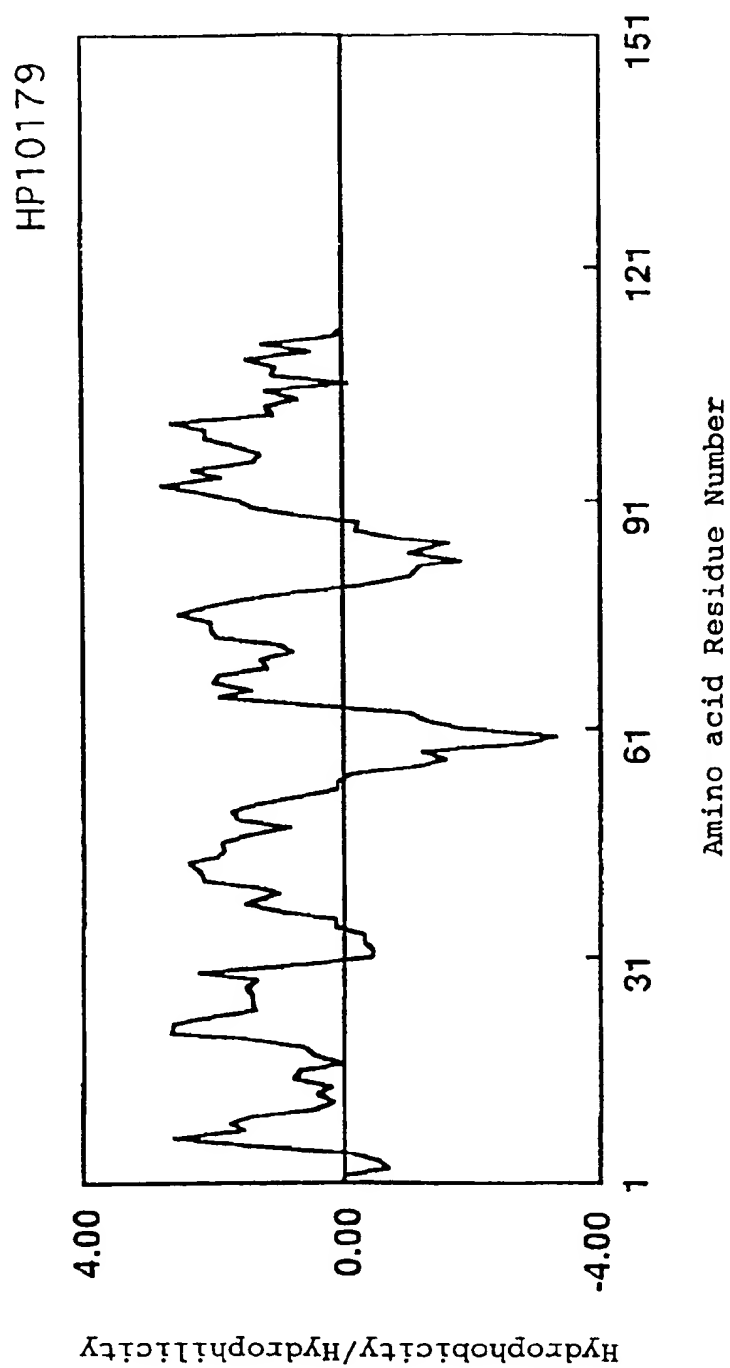
17/28

Fig. 17



18/28

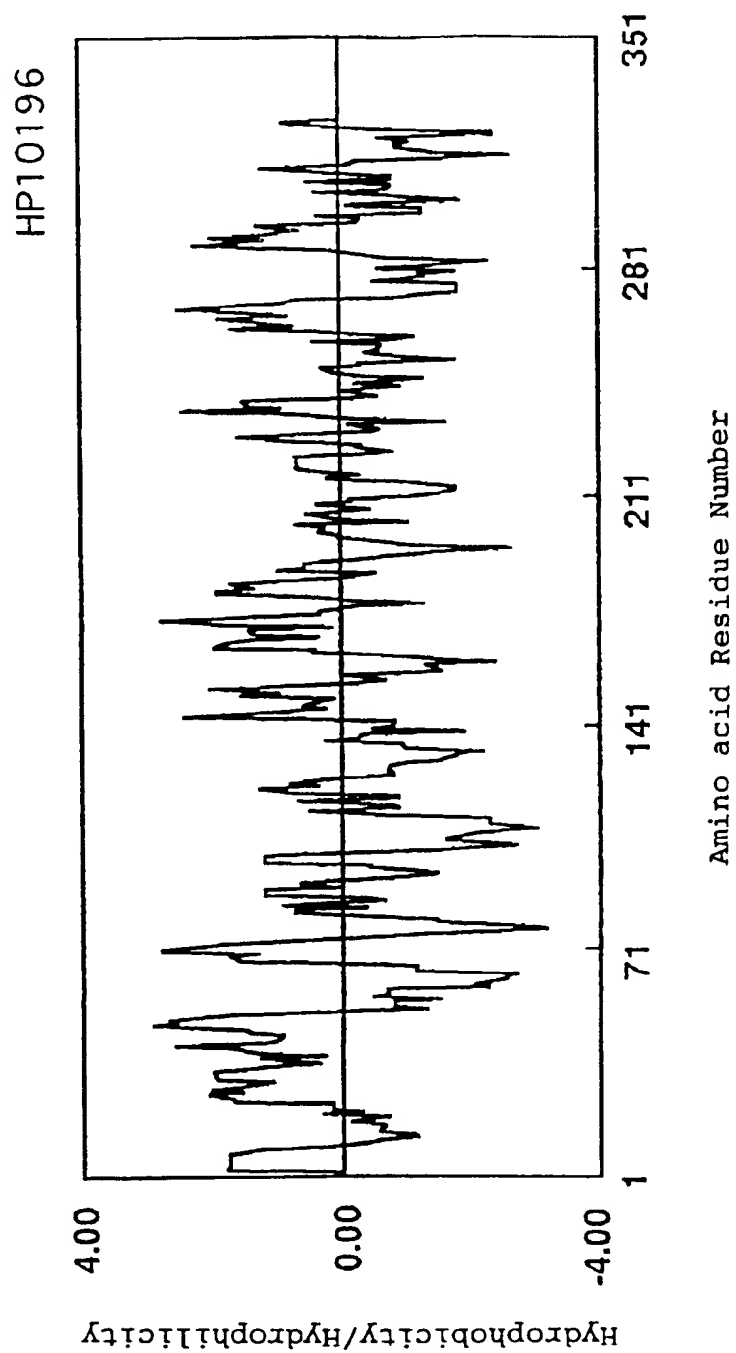
Fig. 18





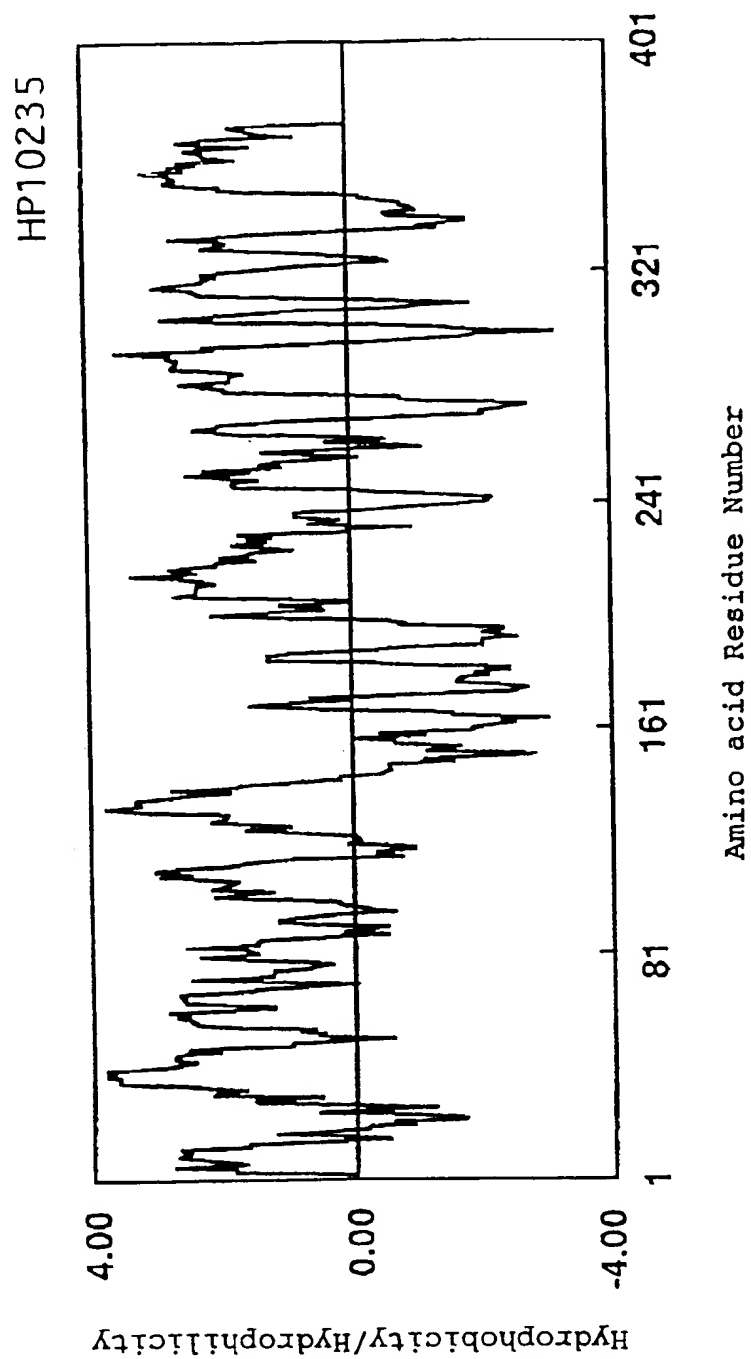
19/28

Fig. 19



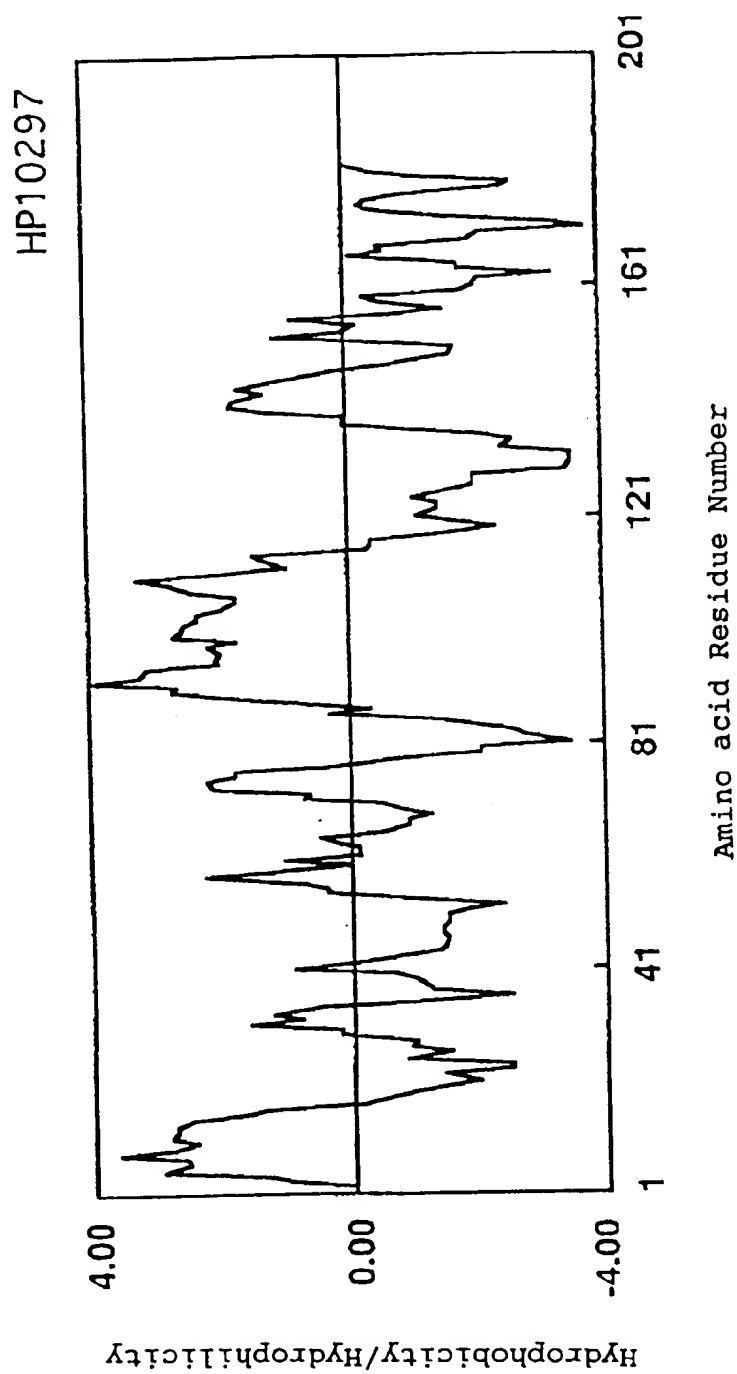
20/28

Fig. 20



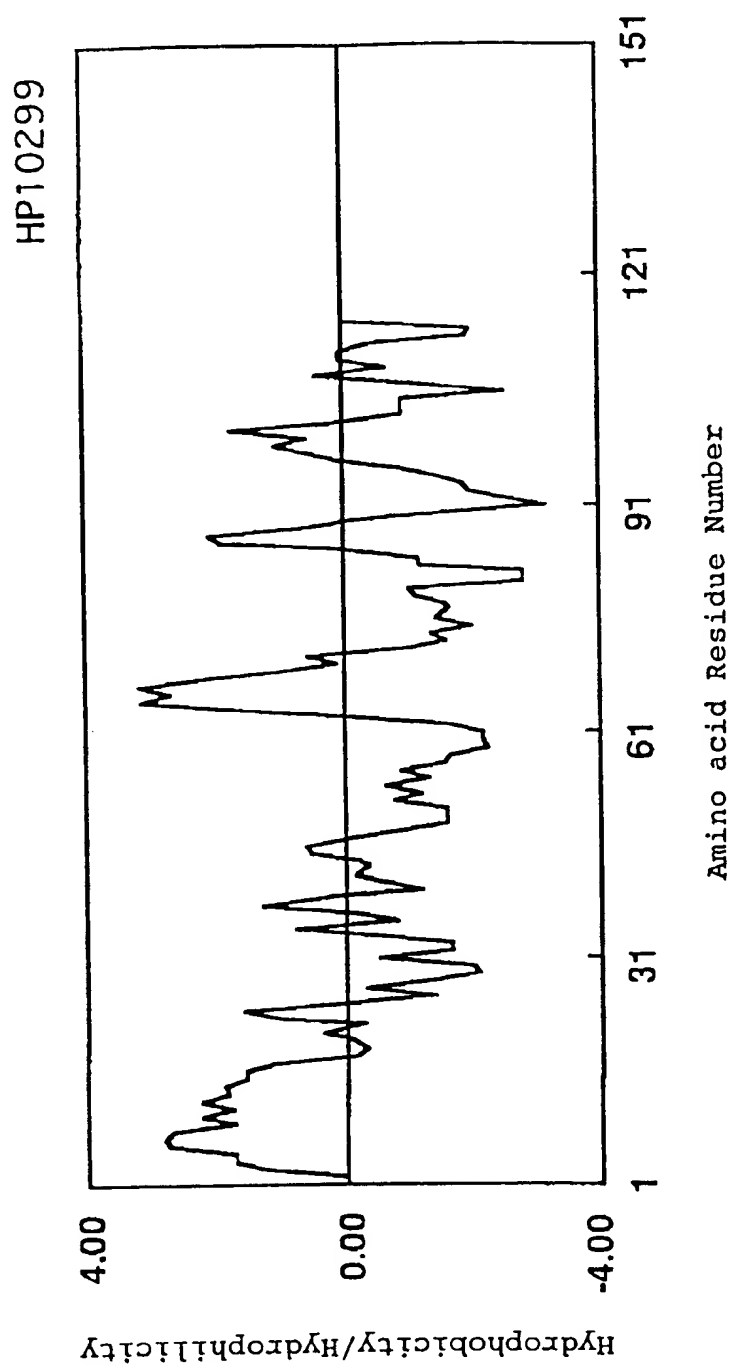
21/28

Fig. 21



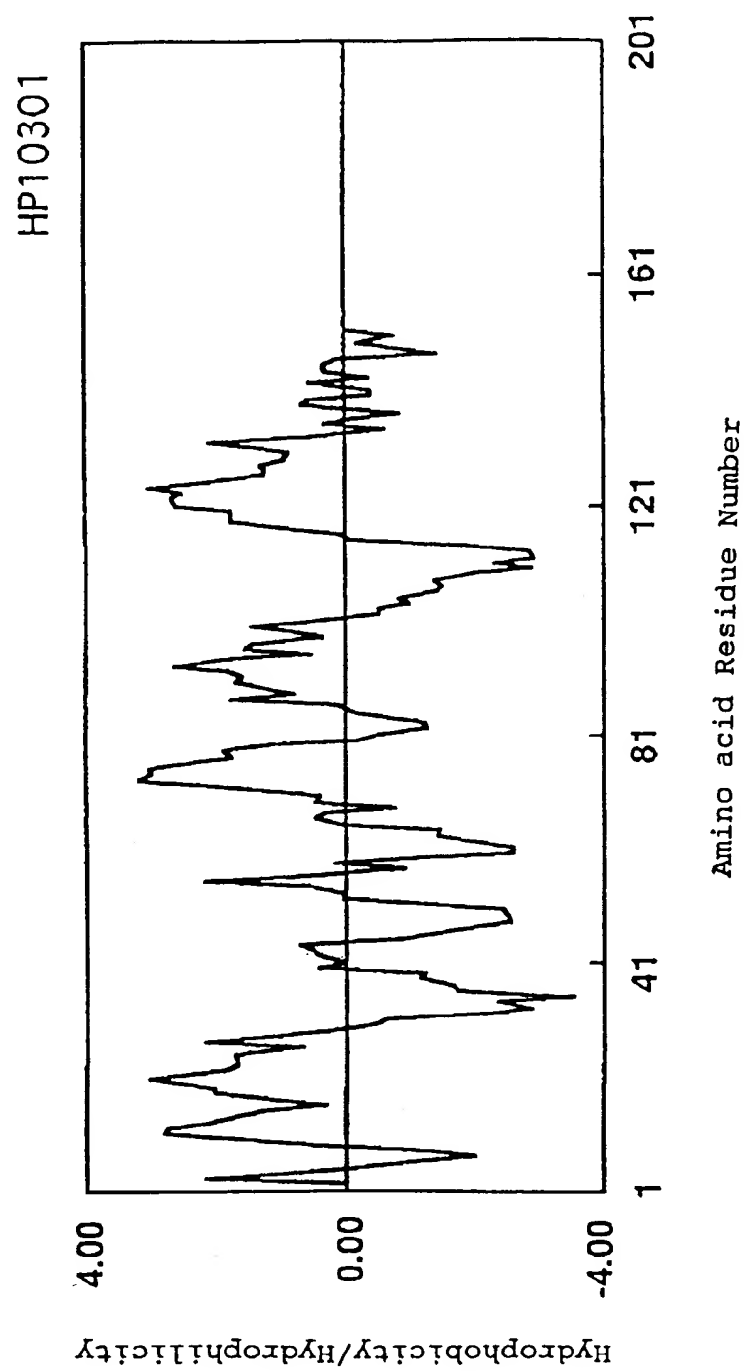
22/28

Fig. 22



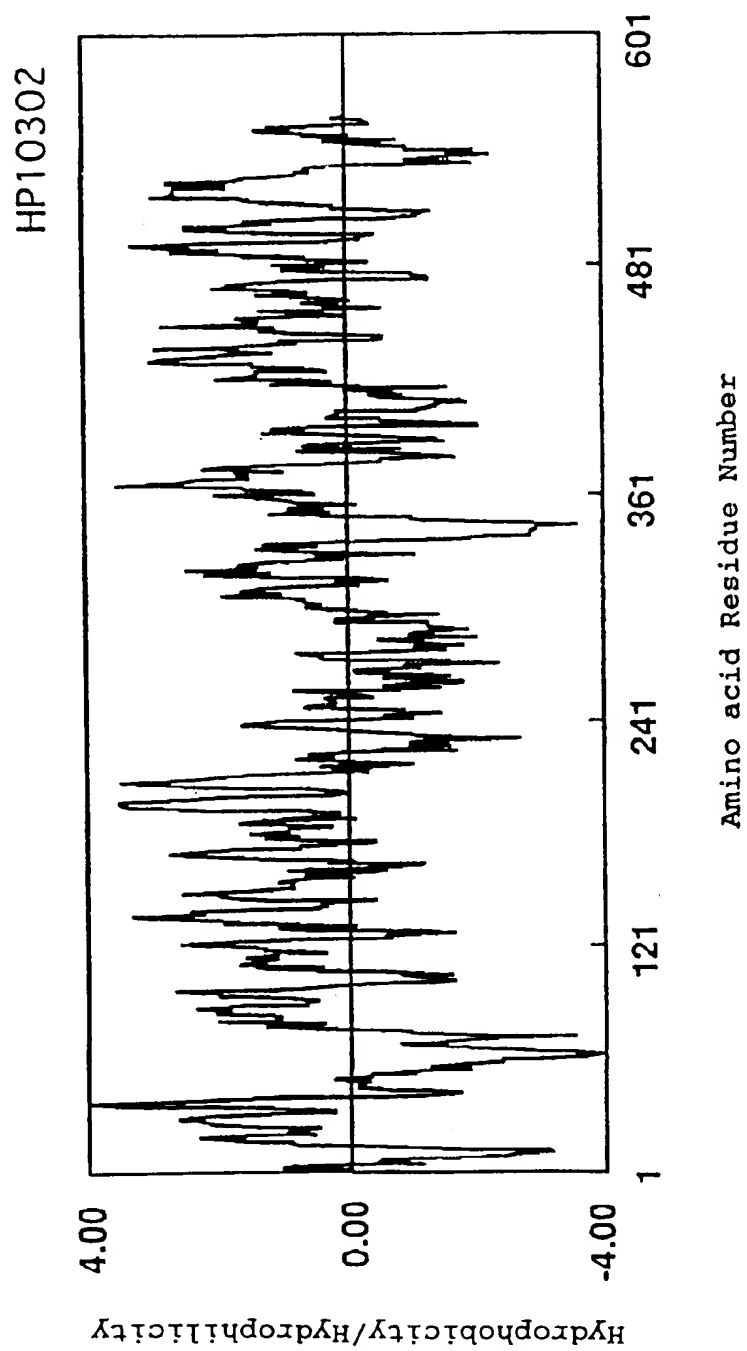
23/28

Fig. 23



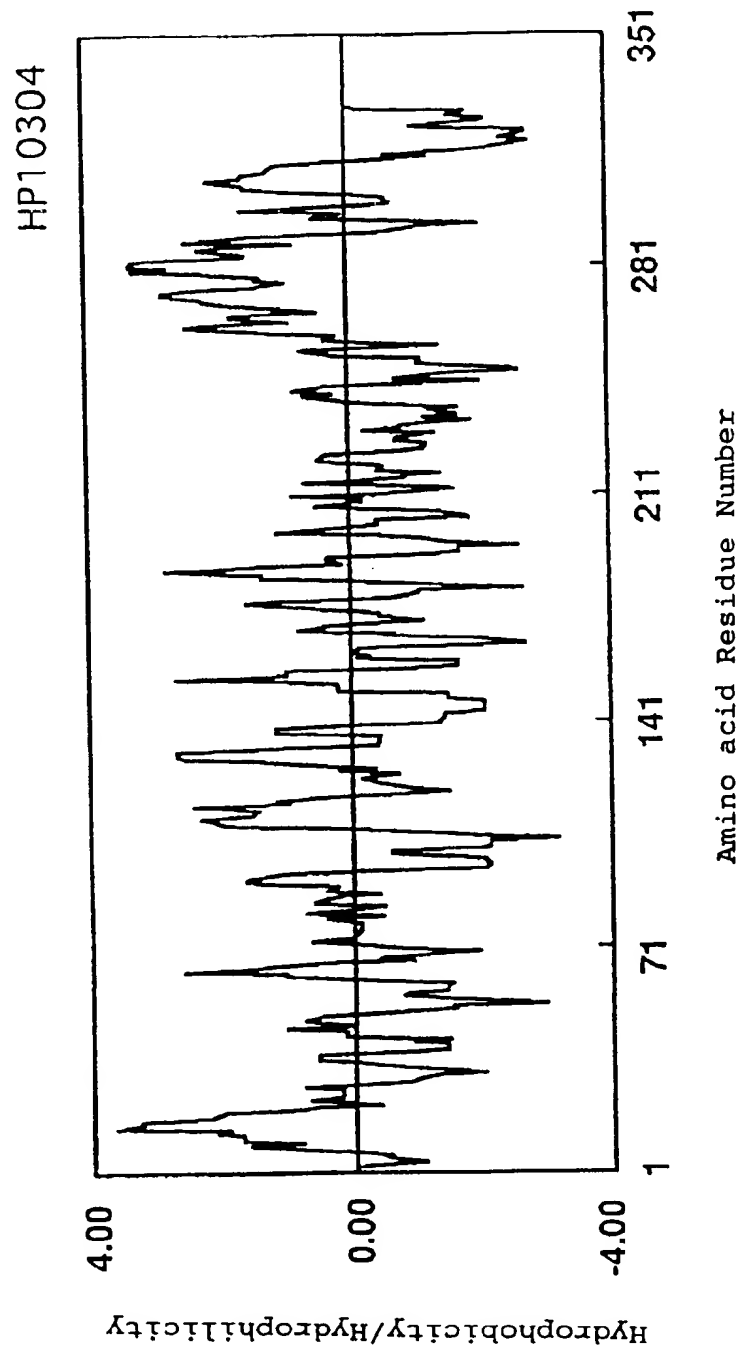
24/28

Fig. 24



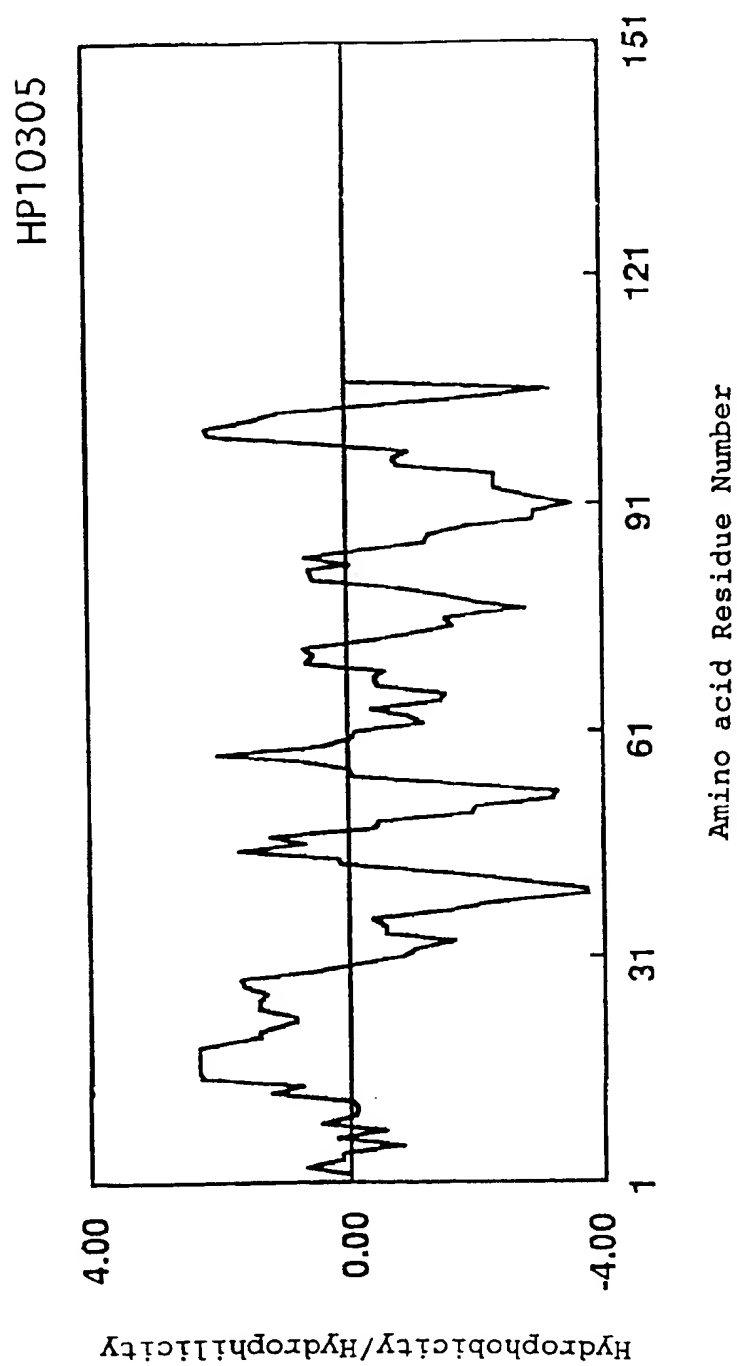
25/28

Fig. 25



26/28

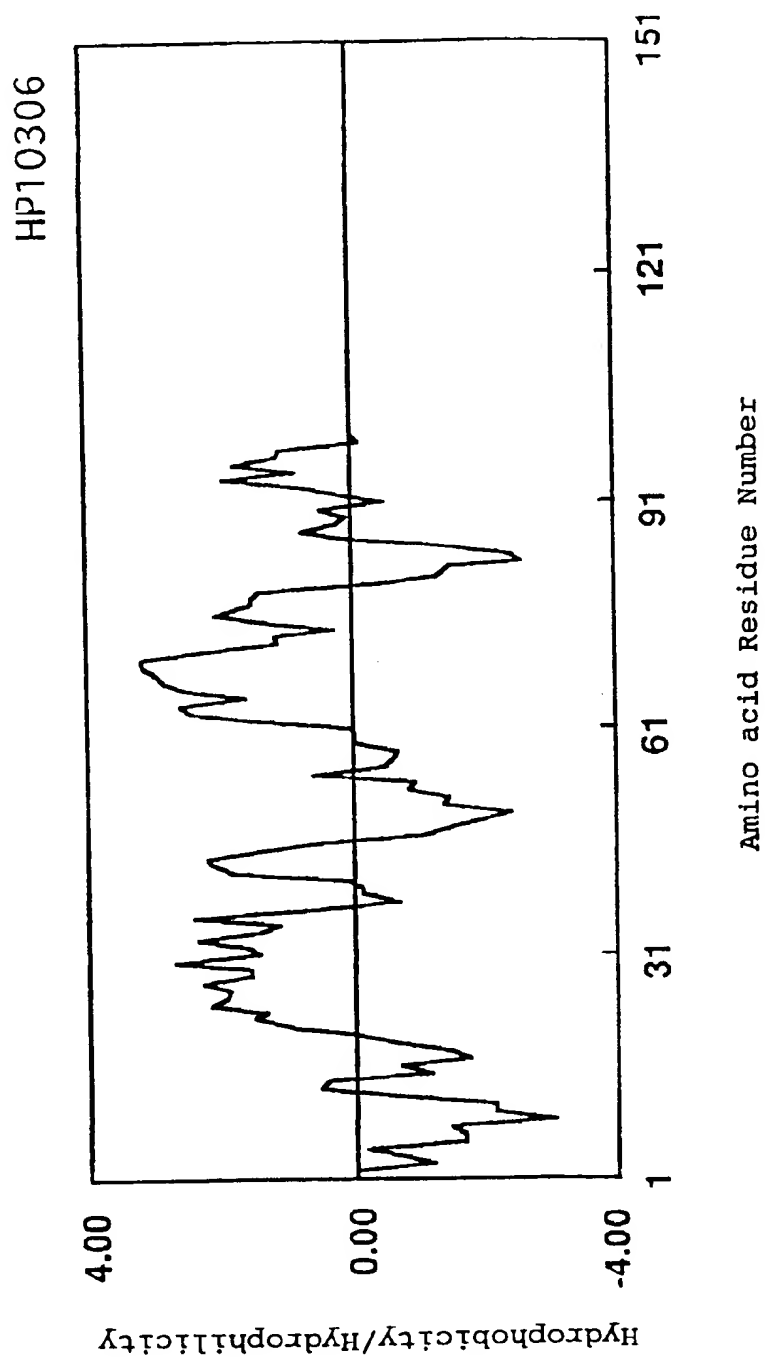
Fig. 26





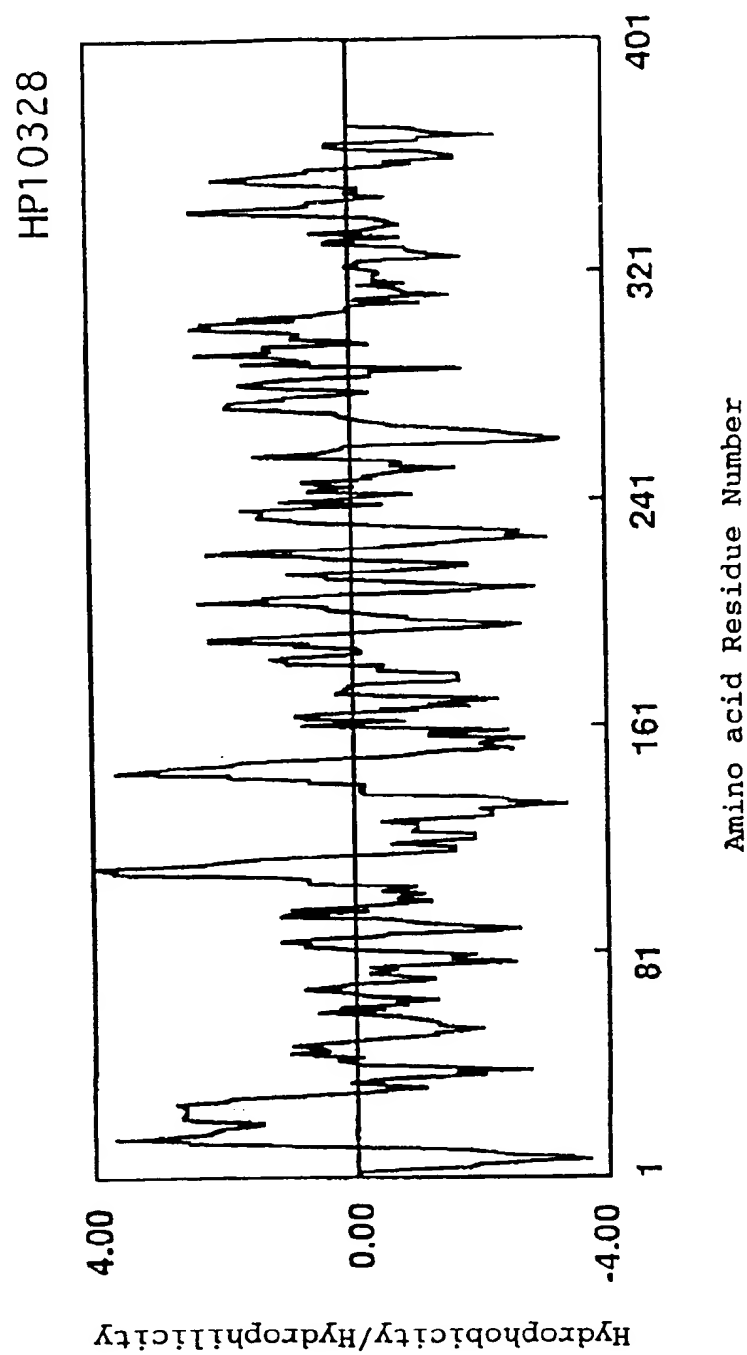
27/28

Fig. 27



28/28

Fig. 28



**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification</b> <sup>6</sup> : <b>C12N 15/12, C07K 14/705, C12N 5/10, 15/57, 9/48, 9/14, 15/55</b>	<b>A3</b>	<b>(11) International Publication Number:</b> <b>WO 98/21328</b> <b>(43) International Publication Date:</b> 22 May 1998 (22.05.98)
<b>(21) International Application Number:</b> PCT/JP97/04056 <b>(22) International Filing Date:</b> 7 November 1997 (07.11.97)  <b>(30) Priority Data:</b> 8/301429 13 November 1996 (13.11.96) JP  <b>(71) Applicants (for all designated States except US):</b> SAGAMI CHEMICAL RESEARCH CENTER [JP/JP]; 4-1, Nishi-Ohnuma 4-chome, Sagamihara-shi, Kanagawa 229 (JP). PROTEGENE INC. [JP/JP]; 2-20-3, Naka-cho, Meguro-ku, Tokyo 153 (JP).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> KATO, Seishi [JP/JP]; 3-46-50, Wakamatsu, Sagamihara-shi, Kanagawa 229 (JP). SEKINE, Shingo [JP/JP]; 4-4-1, Nishi-Ohnuma, Sagami-hara-shi, Kanagawa 229 (JP). YAMAGUCHI, Tomoko [JP/JP]; 5-13-11, Takasago, Katsushika-ku, Tokyo 125 (JP). KOBAYASHI, Midori [JP/JP]; 647-2, Chougo, Fujisawa-shi, Kanagawa 252 (JP).  <b>(74) Agents:</b> AOYAMA, Tamotsu et al.; Aoyama & Partners, IMP Building, 3-7, Shiromi 1-chome, Chuo-ku, Osaka-shi, Osaka 540 (JP).	<b>(81) Designated States:</b> AU, CA, JP, MX, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>  <b>(88) Date of publication of the international search report:</b> 20 August 1998 (20.08.98)	
<b>(54) Title:</b> HUMAN PROTEINS HAVING TRANSMEMBRANE DOMAINS AND DNAs ENCODING THESE PROTEINS  <b>(57) Abstract</b>  Proteins containing any of the amino acid sequences represented by Sequence No. 1 to Sequence No. 2 or by Sequence No. 4 to Sequence No. 25 and DNAs encoding said proteins exemplified by cDNAs containing any of the base sequences represented by Sequence No. 26 to Sequence No. 50. Said proteins can be provided by expressing cDNAs encoding human proteins having transmembrane domains and recombinants of these human cDNAs.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## INTERNATIONAL SEARCH REPORT

International Application No.

PCT/JP 97/04056

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/12 C07K14/705 C12N5/10 C12N15/57 C12N9/48  
 C12N9/14 C12N15/55

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JOURNAL OF MOLECULAR BIOLOGY, vol. 157, no. 1, 5 May 1982, pages 105-132, XP000609692 KYTE J ET AL: "A SIMPLE METHOD FOR DISPLAYING THE HYDROPATHIC CHARACTER OF A PROTEIN" cited in the application see abstract	1-5
Y	SCIENCE, vol. 272, 10 May 1996, pages 872-877, XP002031517 FENG Y ET AL: "HIV-1 ENTER COFACTOR: FUNCTIONAL CDNA CLONING OF A SEVEN-TRANSMEMBRANE G PROTEIN-COUPLED RECEPTOR" cited in the application see the whole document	1-5

☒ Further documents are listed in the continuation of box C.☐ Patent family members are listed in annex.

## \* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search

12 March 1998

Date of mailing of the international search report

03.07.98

Name and mailing address of the ISA

European Patent Office, P.B. 5618 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax: (+31-70) 340-3016

Authorized officer

ESPEN, J

## INTERNATIONAL SEARCH REPORT

Intern: al Application No

PCT 97/04056

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	J BIOL CHEM, APR 12 1996, 271 (15) P8549-52, UNITED STATES, XP002058790 HOLLOWAY MP ET AL: "A hydrophobic domain of Ca <sup>2+</sup> -modulating cyclophilin ligand modulates calcium influx signaling in T lymphocytes." see abstract ---	1-5
Y	BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS., vol. 168, 1990, ORLANDO, FL US, pages 574-579, XP002058791 APPERSON M ET AL: "A yeast protein, homologous to the proteolipid of the chromaffin granule proton-ATPase, is important for cell growth" see figure 2 ---	1-5
P,X	EMHUM1 Database entry HSD052 Accession number D89052; 07 Dec 1996 NISHIGORI H ET AL: 'Cloning and chromosomal localization of the gene encoding a protein homologous to the yeast protein PPA1, an proton-ATPase-like protein' XP002058792 see sequence -----	1-5

# INTERNATIONAL SEARCH REPORT

International application No.

P 97/ 04056

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see continuation-sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
  
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
  
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-5 in part (subject 1. on next sheet)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

## 1. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 26 and 51 and protein relating to  
SEQ ID No 1

## 2. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 27 and 52 and protein relating to  
SEQ ID No 2

## 3. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 28 and 53 and protein relating to  
SEQ ID No 3

## 4. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 29 and 54 and protein relating to  
SEQ ID No 4

## 5. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 30 and 55 and protein relating to  
SEQ ID No 5

## 6. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 31 and 56 and protein relating to  
SEQ ID No 6

## 7. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 32 and 57 and protein relating to  
SEQ ID No 7

## 8. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 33 and 58 and protein relating to  
SEQ ID No 8

## 9. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 34 and 59 and protein relating to  
SEQ ID No 9

## 10. Claims: Claims 1-5 in part



## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

DNAs relating to SEQ ID No 35 and 60 and protein relating to  
SEQ ID No 10

## 11. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 36 and 61 and protein relating to  
SEQ ID No 11

## 12. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 37 and 62 and protein relating to  
SEQ ID No 12

## 13. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 38 and 63 and protein relating to  
SEQ ID No 13

## 14. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 39 and 64 and protein relating to  
SEQ ID No 14

## 16. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 41 and 66 and protein relating to  
SEQ ID No 16

## 17. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 42 and 67 and protein relating to  
SEQ ID No 17

## 18. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 43 and 68 and protein relating to  
SEQ ID No 18

## 19. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 44 and 69 and protein relating to  
SEQ ID No 19

## 20. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 45 and 70 and protein relating to  
SEQ ID No 20

# INTERNATIONAL SEARCH REPORT

International Application No. PCT/ JP 97/04056

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

21. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 46 and 71 and protein relating to  
SEQ ID No 21

22. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 47 and 72 and protein relating to  
SEQ ID No 22

23. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 48 and 73 and protein relating to  
SEQ ID No 23

24. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 49 and 74 and protein relating to  
SEQ ID No 24

25. Claims: Claims 1-5 in part

DNAs relating to SEQ ID No 50 and 75 and protein relating to  
SEQ ID No 25